1979

The assessment of perceptual-motor skills of preschool-age children using norm-referenced and criterion-referenced test characteristics

Robert Wesley Fuqua
Iowa State University

INFORMATION TO USERS

7916195

FUQUA, ROBERT WESLEY
THE ASSESSMENT OF PERCEPTUAL-MOTOR SKILLS OF PRESCHOOL-AGE CHILDREN USING NORM-REFERENCED AND CRITERION-REFERENCED TEST CHARACTERISTICS.

IOWA STATE UNIVERSITY, PH.D., 1979

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

The assessment of perceptual-motor skills of preschool-age children using norm-referenced and criterion-referenced test characteristics

by

Robert Wesley Fuqua

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Psychology

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1979

TABLE OF CONTENTS

	Page
ABSTRACT	iv
INTRODUCTION	1
The Problem	1
Norm-referenced Testing	2
Criterion-referenced Testing	6
Domain-referenced Testing	18
Combining CRT and NRT Characteristics	24
A New Perspective	27
Assessing Motor Development in Early Childhood	31
The Proposal	39
METHODS	41
Subjects	41
Testing Materials	41
Procedures	44
Design	46
RESULTS	47
Normative Data	47
Reliability	74
Inter-correlation among Tasks, Functional Areas and Domains	85
Analyses of Variance	92
CONCLUSIONS	100
Selection and Organization of Content	101
The Scoring System	105
REFERENCE NOTES	112
REFERENCES	114
APPENDIX A: PRESCHOOL MOTOR ASSESSMENT BIBLIOGRAPHY	123
APPENDIX B: PRESCHOOL PERCEPTUAL MOTOR SCALE	129
Scoring Procedures	148
Data Comparison Procedures: A Detailed Example	150
Specific Allocation of Points	153

APPENDIX C: ANALYSES OF VARIANCE TABLES 160

ACKNOWLEDGMENTS 187

ABSTRACT

A study was designed to provide a new perspective for educational measurement. Specifically, a measurement approach combining the strengths of norm-referenced testing (NRT) and criterion-referenced testing (CRT) was suggested. To demonstrate the feasibility of such an approach, a perceptual-motor scale for preschool-age children was designed and administered to a group of three, four, and five year old children. Design characteristics of the test, such as the organization of content and the instructional utility of the scoring system, were investigated. While future research is needed to improve certain aspects of the test, the results generally supported the conclusion that combining CRT and NRT characteristics is a logical and feasible approach to meet current instructional needs.

INTRODUCTION

The Problem

Within the field of education, two major approaches to measuring educational behavior are usually identified. While they are labeled norm-referenced testing (NRT) and criterion-referenced testing (CRT), this terminology has not led to clear distinctions. In fact, many different uses of these labels, especially that of CRT, have resulted in much confusion (Denham, 1975; Hambleton, Swaminathan, Algina, & Coulson, 1978; Glass & Smith, 1978; Hively, 1974; Meskauskas, 1976; Messick, 1975; Millman 1974a; Tallmadge & Horst, 1976; Davis, Note 1).

Some writers (e.g., Carver, 1974; Denham, 1975; Tallmadge & Horst, 1976; Fuqua & Phye, Note 2) believe that some of the confusion surrounding NRT and CRT stems from a tendency to view the two approaches as representing mutually exclusive ends of a continuum, i.e., a newly constructed test must be either norm-referenced or criterion-referenced. Strong rhetoric often has been used in defending or advocating a measurement approach. This rhetoric and the tone of the arguments accompanying it have at times resulted in adversary relationships (e.g., Block, 1971; Ebel, 1971; Hoffman, 1964; Popham, 1974; 1975; 1976).

The thesis of this paper is that while measurement differences have existed between NRT and CRT, a clear separation is not possible and that attempts should be made to construct and use tests which combine both NRT and CRT properties. Furthermore, to demonstrate this possibility, the construction of such an instrument was attempted. In beginning this task, a general description and an evaluation of the two measurement approaches was made. This discussion led to the construction of a device to measure perceptual-motor skills in preschool-age children which combined the separate strengths of each measurement approach while minimizing their separate weaknesses.

Norm-referenced Testing

Origins of NRT in education can be found in Binet's work at the turn of the century in Paris, while the major early American efforts were conducted by J. McKeen Cattell, E. L. Thorndike, Goddard, and Terman (Charles, 1970). Over half a century has passed since these beginnings, providing ample opportunity for improvement, refinement, and dissemination of ideas. As a result, a sophisticated psychometric theory has evolved and received widespread use.

The basic concepts upon which NRT are based are psychometric scaling, latent-trait theory, and correlational techniques. These concepts form an integral part of classical measurement theory and are the basis for determining test reliability and validity, selecting items, and deriving scores.

All of these theoretical ideas continue to be developed. However, a closer look is needed at how these concepts are generally applied.

Reliability and validity are two of the main concepts of classical measurement theory. Reliability studies usually are concerned with the degree of consistency of test scores and the particular causes for variability. Validity studies have focused on what and how well a test measures. Generally, a scale is considered reliable when individuals obtain the same relative scores on alternative forms of a test or over a period of time. Furthermore, a test is usually considered valid when scores on the scale correlate with relative performance in other situations or with other relevant measures.

Classical measurement theories also have developed item selection procedures which are designed to enhance a test's reliability and validity. Items usually are selected for NRT on the basis of their ability to discriminate between those who possess the knowledge or ability being measured and those who do not. Based upon the assumption that the test items are measuring the same thing, discrimination indexes are computed by comparing individual item responses to total test score. Since items which everyone answers correctly or incorrectly do not distinguish among individuals, items with middle difficulty levels (e.g., .40 - .60) are selected most often.

Scores on norm-referenced tests are derived by referring to a table of norms. These scores are interpreted by comparing an individual's performance in relation to other individuals who are representative of a certain population. For example, a student taking a norm-referenced spelling test could be found to score at the 70th percentile with respect to a norm group of 1,000 other students, i.e., 30 percent of these students spelled more words correctly. Frequently scores from norm-referenced achievement tests are used to draw prediction and/or selection inferences about an individual's performance. A student's scores on a spelling test could be used to predict later spelling behavior (prediction) and/or determine if the student should be placed in a special spelling program (selection).

Uses of NRT. NRT has often been used to assist in making educational decisions about individual students. For example, a teacher may refer a student who has generally performed poorly in academic activities for educational and/or psychological evaluation. One or a variety of standardized NRT devices, such as individual achievement or intelligence tests, may be administered to help determine the child's ability level and/or to draw inferences about future performance. Raw scores from these measures are transformed to yield classifications such as descriptive verbal labels (e. g., mildly retarded), percentile ranks or grade equivalent scores. Ysseldyke and Salvia (1974) have described this assessment approach and have called it the "ability-training" model. Often those in education using this approach assume that specific abilities underlie academic skills. Furthermore, these individuals assume that intervention needs to be directed at remediating abilities so that skill performance can improve.

NRT has not only been used to make educational decisions about individuals, but also has been used in the evaluation of groups of students. Standardized norm-referenced tests such as nationally normed achievement tests (e.g., Metropolitan Achievement Test, Iowa Test of Basic Skills) are usually administered to large groups of students once or twice a year to evaluate educational efforts in relation to a representative national sample of students. Other norm-referenced achievement tests are administered to groups of students at the end of a unit or course of study to determine whether the goals of instruction have been accomplished or not.

Criticisms of NRT. There have been many and varied criticisms of NRT. Some critics agree theoretically that NRT can fulfill the purposes for which it has been intended, but believe that these purposes are not often attained. For example, Salvia and Ysseldyke (1978) have criticized the usage of tests which fall below generally acceptable standards of reliability, and also use small, poorly representative samples for norming. While criticisms such as these are justified, they reflect more the abuses of test users and constructors than fundamental inadequacies of NRT. Other critics have become dissatisfied with the limitations of NRT and have proposed alternatives generally identified as CRT (e.g., Glaser, 1963).

Criterion-referenced Testing

Historically, the concept of CRT is not new. However, Glaser usually is credited with introducing the term and the concept of CRT to contemporary educational measurement. Glaser (1963) described a need for a measurement approach which would provide information NRT failed to provide. Particularly, Glaser saw a need for specific behavioral information about an individual's performance on an achievement continuum. He contrasted this with NRT information which describes an individual's relative position within a norm group. Glaser believed that NRT failed to produce the information he wanted because test content was defined generally, rather than precisely. Furthermore, he believed the dimensions assumed to underlie NRT scores were vaguely understood.

> Underlying the concept of achievement measurement is the notion of a continuum of knowledge acquisition ranging from no proficiency at all to perfect performance. An individual's achievement level falls at some point on this continuum as indicated by the behaviors he displays during testing. The degree to which his achievement resembles desired performance at any specified level is assessed by criterion-referenced measures of achievement or

> proficiency. The standard against which a student's performance is compared when measured in this manner is the behavior which defines each point along the achievement continuum. The point is that the specific behaviors implied at each level of proficiency can be identified and used to describe the specific tasks a student must be capable of performing before he achieves one of these knowledge levels (Glaser, 1963, pp. 519-520).

Unfortunately, the evolution of CRT has not produced the results Glaser hoped for. In fact, as Glass and Smith (1978) have commented, the development of CRT has been "...a case study in confusion and corruption of meaning" (p. 13). This confusion can be seen (1) in the misinterpretation of the term criterion, (2) in the use of cut-off scores with CRT, (3) in the various definitions of CRT, and (4) in the association of specific behavioral objectives with CRT.

CRT and the term criterion. While Glaser (1963) did at times discuss the term "criterion" as a performance standard and intimated the use of cut-off scores between competence and incompetence, he repeatedly emphasized "...a continuum of knowledge acquisition ranging from no proficiency at all to perfect performance" (p. 519). He also emphasized a "...degree of competence" (p. 520). For example, in 1963 Glaser wrote:

> The standard against which a student's performance is compared in order to obtain the first kind of information is the criterion behavior which defines increasing subject matter competence along a continuum of achievement (p. 800).

In 1971, Glaser and Nitko further clarified their use of performance standards by stating:

> A criterion referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards. Performance standards are generally specified by defining a class or domain of tasks that should be performed by the individual. Measurements are taken on representative samples of tasks drawn from this domain, and such measurements are referenced directly to this domain for each individual measured (Glaser & Nitko, 1971, p. 653).

Glaser (Glaser & Nitko, 1971) cautioned against confusing the term "criterion" with a minimum proficiency level.

> A second prevalent interpretation of the term criterion in achievement measurement concerns the imposition of an acceptable score magnitude as an index of attainment. The phrases "working to criterion level" and "mastery is indicated by obtaining ascore equivalent to 80 percent of the items correct" are indicative of this type of criterion (p. 653).

In further refining the meaning of CRT, Glaser (Glaser & Nitko, 1971) distinguished CRT from NRT by stating that CRT devices were not designed to:

> ...facilitate individual difference comparisons as relative standing of an examinee in a norm group or population, nor are they designed to facilitate interpretations about an examinee's relative standing with respect to a hypothetical variable such as reading ability (p. 653).

It is clear from these excerpts of Glaser's writings that he was interested in the construction of measurement devices which provided information about an individual's performance relative to a specified domain of tasks and not solely to some mastery-nonmastery level of performance. Despite Glaser's efforts, the term criterion became synonymous with performance standard.

Popham (e.g., Popham & Husek, 1969; Millman & Popham, 1974; Popham, 1975) contributed to the interpretation of the term criterion as a specific performance standard. For example, Popham and Husek (1969) wrote:

> Criterion-referenced measures are those which are used to ascertain an individual's status with respect to some criterion, i.e., performance standard (Popham & Husek, 1969, p. 2).

Later in the same article, Popham and Husek discussed the interpretation of a CRT score in terms of a minimum level of competency:

> For example, suppose that an instructional objective had been devised which required a learner to multiply correctly pairs of three digit numbers. We would prepare 20 items composed of randomly selected digits to measure this skill. Because of possible computation errors, the required proficiency level for each successful student might be set at 90 percent, or better, thereby allowing errors on two of the 20 items. In reporting an individual's performance on a test such as this, one alternative is to once more use an "on-off" approach, namely, either the 90 percent minimum has been achieved or it hasn't (p. 7).

Popham (e.g., Popham & Husek, 1969; Millman & Popham, 1974) also interpreted the term criterion as a performance standard when he discussed score variability in CRT.

> With criterion-referenced tests, variability is irrelevant ...variability is not a necessary condition for a good criterion referenced test (Popham & Husek, 1969, p. 3).

In support of this viewpoint, Popham argued that it is possible for everyone to miss all items before instruction, and to get all items correct after instruction. However, as Woodson (1974b) has pointed out, this would be the case when the outcome of CRT is viewed "...as an all or nothing decision, the classifying of an individual's knowledge as nonmastery or mastery" (p. 140).

Popham and Husek's (1969) article is one of the most cited papers on CRT and has done a lot to perpetuate the cut-off, mastery-nonmastery interpretation of the term criterion. However, others such as John Carroll (1970) also contributed to this confusion. Because of these and similar writings the term criterion has not become associated with a continuum of behavior as originally intended.

CRT and cut-off scores. The misinterpretation of criterion led to research on the use of cut-off scores, further compounding the confusion surrounding CRT (e.g., Carver, 1970; Hambleton & Novick, 1973; Huynh, 1976; Livingston, 1972; Meskauskas, 1976; Swaminathan, Hambleton & Algina, 1973, 1974; Huyn, Note 3). This research was based on the assumption that the primary purpose of CRT was to classify individuals into mastery/nonmastery states or categories using cut-off scores. Therefore, several of the studies attempted to derive reliability indexes for tests using cut-off scores. Other studies were concerned with particular techniques for establishing cut-off scores.

Recently, Glass and Smith (1978) critically reviewed the different methods of setting mastery scores. Of the six techniques identified by them, all were criticized for being either "...blatantly arbitrary or derived from a set of arbitrary premises..." (p. 17). In addition, Glass and Smith specifically criticized two techniques because they produced scores not referenced directly to tasks, but to a norm group. In other words, scoring techniques which had been described as criterion-referenced were in fact norm-referenced. While Glass and Smith were not the first to reach these conclusions (e.g., Millman, 1973; Subkoviak & Baker, 1977), they were the first to provide a comprehensive critique of the area.

Glass and Smith's (1978) review clearly demonstrated that many of the procedures used to establish cut-off scores heavily rely upon subjective judgemental procedures which have resulted in inconsistent and faulty decisions. For example, one of these techniques has experts study tests and test items and judge what score should be achieved by a "minimally" competent individual. Andrews and Hecht (1976) empirically compared two alternative refinements of this method. Results of their experiment revealed a 23 percent gap between what judges using one method and judges using an alternative method considered the mimimal percentage of passed items needed for competence. Glass and Smith (1978) concluded:

> The attempt to base criterion scores on a concept of minimal competence fails for two reasons: (1) it has virtually no foundation in psychology of learning...; (2) when its arbitrariness is granted but judges attempt nonetheless to specify minimal competence, they disagree widely. In short, the idea of minimal competence is bad logic and even worse psychology (p. 16).

While Glass and Smith (1978) identified two of the six standard setting techniques as more norm referenced than criterion referenced, and three others as subjective and arbitrary, the last one was described as an objective attempt to establish criterion scores. Glass and Smith labeled this method as the "Operations Research Method". Block (Note 4) who originally presented this method, theoretically argued that a relationship should exist between performance on a transfer test, or any other valued outcome, and degree of mastery on a criterion-referenced test. This relationship is drawn on a graph to determine the level of performance on the criterion-referenced test for which performance on the transfer test is maximized. While this method is objective, it doesn't satisfactorially determine the cut-off score unless the relationship between CRT performance and the valued outcome is nonmonotonic. A monotonic function would set the criterion score at 100 percent, one with which no measurement error could exist. Obviously, this is an unrealistic standard.

Block (Note 4) failed in an attempt to empirically demonstrate the efficacy of the "Operations Research Method". When the outcome variables were plotted for four groups of subjects who had acquired different levels of mastery, the function was monotonic increasing and the F-ratio for nonlinearity nonsignificant. As a result, a nonarbitrary or "rationally" defensible criterion could not be established.

In summary, the various attempts to establish absolute standards of performance have not produced clear, consistent measurement procedures needed for the instructional purposes identified by Glaser (1963). In fact, Glass and Smith (1978) stated that confusion in this area comes from the belief that absolute standards could be established. They concluded:

> ...In education there are virtually no absolute standards of value. "Goodness" and "badness" must be replaced by the essentially comparative concepts of "better" and "worse".... Considerable clarity and consensus are bought when "change" is substituted for "absolute level of performance", even if all problems

are not solved (p. 17).

CRT definitions. While much confusion surrounded the misuse of the term criterion, additional confusion resulted from definitions of CRT that exclusively emphasized one or two aspects of CRT. For example, Kriewall (Note 5) chose to emphasize homogeneity of item content. Specifically, Kriewall required that the tests be homogeneous in difficulty so that the probability of passing or failing any one item was the same for all items. Ivens (Note 6), on the other hand, emphasized behavioral objectives. According to Ivens, a test was not criterion-referenced unless the items were keyed to a set of behavioral objectives. Since Ivens' definition did not include a fixed criterion, could a test be "norm-referenced" by Popham and Husek's (1969) definition and "criterion-referenced" by Ivens' definition? To further the confusion, it seems that a test which included items keyed to behavioral objectives as well as including a specified performance standard, may be "criterion-referenced" according to Popham and Husek's and Ivens' definitions, but would not satisfy Kriewall's definition unless all items possessed equal difficulty indexes.

The proliferation of definitions has become so wide spread that recently Gray (1978) was able to classify 57 descriptions of CRT. Unfortunately, instead of offering a new perspective of educational measurement, Gray provided the fifty-eighth definition of CRT.

While much confusion has surrounded attempts at defining CRT, one element common to all definitions can be found. That element is the emphasis on interpretation of test outcomes in terms of specific behaviors. Unfortunately, most research in CRT has not focused on this commonality, and has instead focused on the use of cut-off scores as indicators of performance standards.

CRT and behavior objectives. The last source of confusion surrounding CRT to be considered is the use of behavioral objectives. About the same time that Glaser's original article appeared, Mager (1962) associated performance standards with behavioral objectives:

> If we can specify at least the minimum performance for each objective, we will have a performance standard against which to test our instructional programs (p. 44, emphasis added).

Popham (1973) furthered the connection of performance standards with behavioral objectives when he specified the establishment of performance standards when writing objectives. Popham and Husek (1969) also emphasized the use of behavioral objectives in the construction of criterion-referenced test items. The usual procedure is to take an objective and attempt to produce items that meet the specifications of the objective. For example, an objective which specified that the student select the response from among alternative responses would probably have a corresponding multiple-choice test item. As a result, usually each item on a criterion-referenced test is keyed to a specific objective.

Mager's (1962) work is perhaps best representative of behavioral objectives. Mager argued that the effectiveness of educational procedures could be increased and that useless, unrelated instruction could be decreased, through the use of behavioral objectives. To further these goals, Mager specified the qualities, characteristics, and conditions associated with the "useful" behavioral objective (Mager, 1962).

Initially, Mager's suggestions were enthusiastically accepted and behavioral objectives were widely used. However, Popham (1975) best described the longer range reaction of educators.

> At one time, early in the sixties, proponents of measurable goals thought that the more specific an objective was, the more instructionally helpful it would be. But we ended up with so many specific objectives we were drowning in them. What a way to go - choked to death by a sea of "The student will list..." and "Given ten true-false items, the learner will...."
>
> It was soon recognized that objectives could be written too specifically to be practically useful (p. 71).

Other criticism of detailed behavioral objectives centered on their limited usefulness. Ebel (1973) provided a good example of this criticism:

> ...there are only a few of the most elementary fields of study in which achievement consists of memorizing a limited number of discrete specific facts like the 100 basic facts of addition, the correct spelling of commonly used words, the multiplication tables, etc. Most achievements in learning call for much more than simple recall.... In such fields of study no list of discrete objectives, however long, can adequately portray the totality of learning achievement in any area (pp. 215-216).

In the early sixties and seventies, criticism of behavioral objectives usually came from outside the CRT/instructional technology movement. However, more recently individuals within CRT (e.g., Baker, 1974; Hively, Patterson & Page, 1968; Hively, 1974; Hively, Maxwell, Rabehl, Sension, & Lundin, 1973; Millman, 1973; 1974a; Nitko & Hsu, 1974; Popham, 1974; 1975) have acknowledged the limitations associated with the use of behavioral objectives. For example, Baker (1974) stated that the behavioral objective approach was overemphasized and had generated a false impression of precision. She particularly believed this was the case when educators dealt with other than the elementary areas of subject matter.

> Because objectives are stated in operational language, they appear to be more teachable. An objective such as "Given a lyric poem, the student will be able to write a 450-word essay on the theme and tone", may look achieveable because it follows the much exalted formula: "Given ...the student will be able to...," but such is not the case.... As long as a "behavioral" verb has been supplied, many consultants and supervisors have little to criticize. Even the more sophisticated may seek only to assign the objective to a higher level category

> on the <u>Taxonomy of Educational Objectives</u> (Bloom et al., 1972) (pp. 17-18).

Consequently, educators have been faced with two options. Either they could be inundated with an infinite number of highly specific behavioral objectives which dealt with trivial behaviors, or they could be frustrated with behavioral objectives which provided insufficient cues regarding what class of behaviors were relevant when remediation was needed. It was difficult to determine what specific group of behaviors related to the objective. Whatever the teacher's approach, the probability of attending to irrelevant aspects of behavior was high, since item content was so often broadly defined (Baker, 1974).

<u>Domain-referenced Testing</u>

Hively (1974) and others (e.g., Duncan, 1974; Hentschke & Levine, 1974; Johnson, 1974; Miller, 1974; Millman, 1973; 1974a; 1974b; Nitko & Hsu, 1974; Popham, 1972; 1974; 1975; Sension & Rabehl, 1974; Johnson, Note 7) have agreed with Baker's (1974) viewpoint and have proposed the use of domains or "amplified objectives" as a compromise between vagueness and overprecision. Hively (1974) has further suggested the replacement of the term criterion-referenced test with the term domain-referenced test (DRT). While the majority of these writers view DRT as a refinement of CRT, they do not universally agree that

that the term DRT should be used. However, Hively (1974) believes that the use of the term DRT will not only help resolve the problem surrounding the use of specific behavioral objectives, but also will reduce the confusion and misinterpretation associated with the use of the term criterion.

Hively (1974) retained the major characteristics of CRT which were originally proposed by Glaser (1963). Hively's (1974) objective was to construct a test that would provide information regarding an individual's performance within a clearly specified domain or class of behaviors. He wanted the boundaries of these domains of behavior clearly stated so that many parallel tests could be produced by systematically sampling from that domain. Hively felt this approach would benefit students by allowing the practicing and testing of similar examples from day to day. Hively also saw this association of instruction and testing as the main distinction between DRT and NRT.

> DRT has something NRT does not, and it is something that is tremendously important to the learner. This extra something may not seem important if the major purpose of testing is to assign children to groups for instruction, to select young people for employment, or to predict their future (relative) success in higher education. NRT is excellent for these purposes. But if the purpose is to keep track of day-to-day progress and to study the conditions that facilitate or inhibit it, then NRT, as it is universally practiced in American schools, is useless (Hively, 1974, p. 8).

As mentioned earlier, not all individuals wanting to improve CRT agree on what aspects of CRT need to be emphasized or which need to be ignored. The central distinction usually made between CRT and DRT is the specification and definition of content rules or limits for domains. However, Popham (1972) saw these domains as "amplified objectives" which provided the direction and clarity of behavioral objectives as described by Mager (1962), but without the overprecision or vagueness that often resulted from their use. Somewhat differently, Baker (1974) described domains as subsets of transferable skills and contrasted them with the CRT objective approach by emphasizing the characteristic of generalizability. She used an illustration:

> The ability to list three causes for the depression could only be a suitable objective if, in domain context, it were modified to concern the generalizable causes of economic decline, of which the 1929 depression was only one example (p. 20).

Baker (1974) saw DRT helping the teacher determine what behaviors were of concern. Furthermore, she saw this testing approach as allowing the educator to teach at a transfer rather than at a rote level.

While many people see DRT as a promising area of research, it is obvious some disagreement exists. Another area in DRT which has generated some differences concerns the process of constructing a domain. While Baker (1974) takes a deductive approach by starting with objectives and then building domains which represent them, Hively takes an inductive approach by expanding and generalizing typical tasks and then supplying the descriptive name to the domains.

Reliability in CRT. A review of the literature on CRT/DRT reliability provided an opportunity to more clearly examine the confusion surrounding the recent developments of CRT. Because many aspects of test construction and use must be considered when test reliability is at issue, each characteristic of the test can be examined in relation to a central concept. Thus, a framework was provided for understanding the differences existing within CRT/DRT.

Popham and Husek (1969) were perhaps the first major proponents of CRT to discuss reliability of CRT. They argued that item and test variance was not a necessary characteristic of CRT, but that they were necessary for NRT. For this reason, they stated that traditional reliability methods were not applicable to CRT. Several other CRT proponents adopted this viewpoint. For example, writers such as Carver (1970), Hambleton and Novick (1973), Huynh (Note 3), and Swaminathan; Hambleton and Algina (1974) suggested that reliability for CRT should be seen as consistency of decisions across repeated testings. In fact, Swaminathan, Hambleton and Algina (1974) suggested the use of <u>coefficient kappa</u> (Cohen, 1968) as such an index, and Huynh (Note 3), suggested further refinements of this index. Unfortunately, all these CRT indexes were based on the use of a passing or criterion score.

Woodson (1974a) recognized this flaw of basing reliability indexes on cut-off scores and challenged Popham and Husek's (1969) arguments. He stated that all test and items "must have variance within the range of interest for which they are calibrated in order to provide useful information" (p. 2).

Millman and Popham (1974) responded to Woodson's (1974a) arguments by providing an instructional example. They described an ideal situation as one where before instruction every student missed all items and after instruction every student got all items correct. Millman and Popham (1974) described this situation as showing no item variance. They further suggested that item and test variance would be more characteristic of NRT than CRT.

Woodson (1974a; 1974b; Note 8) helped clarify the issue of CRT reliability by pointing out that both CRT and NRT items and tests must have variance. Woodson (1974b) demonstrated that an item with no variance provides no information. Since no variance would indicate that all the observations within the range of interest were the same, everyone whose ability was within that range would receive identical scores. In this situation, the educator would have no useful information.

Woodson identified the fallacy in Millman and Popham's (1974) argument by demonstrating that within a particular sample, "...an item or test may have no variance, but in the population of observations from which a test was designed... both items and tests must have variance" (p. 63). Woodson (1974b) further argued that to view the situation otherwise, was to view the outcome of CRT as the classification of individuals into mastery or nonmastery states.

Woodson (1974a) was not opposed to the purposes of CRT. In fact, he agreed with CRT advocates that there was a need for measures which would provide information about where an individual is on a dimension of learning. However, Woodson recognized the problems associated with the use of cut-off scores and suggested a reconceptualization of CRT measurement as it applied to scores and their reliability estimates. In particular, Woodson carefully stipulated that a specific "range of interest" (a population of observations) needed to be specified for each testing situation. For CRT the "range of interest" referred to the behavioral domain being measured; in NRT the "range of interest" was the distribution of the characteristic in some specified population. In other words, the population of reference, or as Woodson (1974a; 1974b) referred to it, the "range of interest", was what distinguishes NRT from CRT; not item or test variance.

Woodson is not alone in his views. Ebel (1973) suggested earlier that conventional test statistics were appropriate to CRT. According to Ebel (1973), the usual methods of computing test reliability and item discrimination were appropriate with either CRT or NRT when scores were used to determine how well an individual has succeeded in a particular program. More recently, Denham (1975) reiterated this viewpoint. She argued that items on a test, regardless of whether they were CRT, DRT, or NRT, must discriminate between those who possess the particular skill or knowledge and those who do not. Denham contended that a lack of positive discrimination indicated that the item was either poorly written or that a lack of variance in the performance of examinees existed. In either case, Denham concluded, the item had failed the main purpose of testing, i.e., to determine who did and who did not possess the skill or knowledge.

Combining CRT and NRT Characteristics

While most of the educational measurement literature over the past 15 years has generally reflected the advocacy of one measurement approach to the exclusion of the other, recent suggestion to combine CRT and NRT characteristics have surfaced. Cronbach, Rajaratnam and Gleser (1963) were pioneers in these efforts. They discussed the possibility of drawing samples of behaviors from concrete domains in such a way as to create norm-referenced scales of performance over the domain. More recently Denham (1975) proposed a parallax view of NRT, CRT and DRT, suggesting that these three measurement approaches are not separate, independent phenomena.

Others have proposed similar viewpoints. Reviewing testing development over the past decade, Carver (1974) suggested that too much effort had been focused on viewing CRT and NRT as separate entities and hoped that future efforts might take a different appraoch.

> It is hoped that future tests will be developed and evaluated with an appreciation of both dimensions so that researchers and practitioners will have better tests with which to solve measurement problems (p. 518).

Even Hively (1974) who has devoted much effort to distinguishing between NRT and DRT, has suggested adding normative references to DRT.

> Domain-referenced measures of performance are currently not easy to communicate. Even though the assumption may be spurious, many people seem to think they understand what is meant by "third grade reading level" better than "ability to read aloud from a random sample of editorial pages at a

> hundred words per minute without skipping and with fewer than five mispronunciations." In order to give the latter measure appeal, we will have to anchor them to the behavior of prototypical persons or groups (p. 145).

Finally, Popham (1976) and Glaser (1976), the two original main proponents of CRT, have encouraged the use of normative data for CRT. Discussing this possibility, Popham (1976) pointed out that criterion-referenced tests can clearly indicate _what_ students can do, but they have difficulty communicating _how well_ the students _should_ be doing. In answer to this problem both Popham (1976) and Glaser (1976) suggested maximizing the utility of CRT through the use of normative data.

A few examples of tests combining CRT and NRT characteristics already exist. While they are somewhat limited in range of subject areas and age groups, they do provide some evidence that efforts in combining CRT and NRT characteristics can be beneficial. Perhaps the largest undertaking of this kind is the _Stanford Achievement Test_ (SAT) (Madden, Gardner, Rudman, Karlsen, & Mervin, 1973). This test is currently being used nationally and has received excellent reviews (e.g., Salvia & Ysseldyke, 1978). The SAT is constructed to measure skill development in several academic areas. This test has two manuals which are concerned with CRT characteristics. One manual has arranged items by major instructional objectives, while the other manual provides instructional objectives and suggestions on how to design instruction to obtain these objectives.

A more recent effort to combine NRT and CRT characteristics is the Stanford Diagnostic Reading Test (SDRT) (Karlsen, Madden & Gardner, 1977). The SDRT was designed to measure a pupil's performance relative to a norm group and to identify individual student strengths and weaknesses in specific reading skills. The Stanford Diagnostic Mathematics Test (SDMT) (Beatty, Madden, Gardner, & Karlsen, in press) is similarly constructed. All of these devices are well-standardized and have demonstrated sufficient scoring reliability. For a complete discussion of these and other similar instruments see Salvia and Ysseldyke (1978).

A New Perspective

From this review several conclusions can be reached. First, prior to the 1960's educational testing theory and practice emphasized methods designed to predict future student performance or to select students for specific programming. Generally, NRT was quite successful in fulfilling these purposes. However, the emergence of new instructional technology brought requests for measurement procedures which could be more beneficial to day-to-day instructional programming. In response to these requests a measurement approach called criterion-referenced testing (CRT) surfaced.

Early proponents of CRT, such as Glaser (1963), emphasized the construction of criterion-referenced tests which would provide measurements referenced directly to a domain of tasks. Glaser further emphasized that CRT should not be designed to provide individual difference comparisons nor to determine an individual's relative standing with respect to a latent trait such as math ability.

Despite these early efforts to define CRT, confusion surrounded its development. This confusion stemmed from how CRT should be defined and the misinterpretation of the term criterion in CRT. The misunderstanding of "criterion" led to the use of cut-off scores and the development of corresponding scoring and reliability techniques. In addition to these problems, a growing consensus developed among measurement and instructional specialists that behavioral objectives had been wrongly used in writing CRT tests.

Because of the problems surrounding the evolution of CRT, the educational needs identified by Glaser (1963) and others have not been met. Meanwhile, educators still need a system to monitor the day-to-day progress of their students and to study the conditions that effect it. Unfortunately, neither NRT nor CRT are successfully meeting these needs.

Recent attempts have been made to clarify the problems existing within CRT and to provide the measurement techniques required by today's educator. Some individuals have suggested that domains or "amplified objectives" and domain-referenced testing (DRT) can supply the answers. Others have suggested the combining of NRT and CRT characteristics to provide new measurement alternatives. However, both suggestions have limitations.

The problems associated with CRT cannot be eliminated by combining them with NRT characteristics in this fashion. These attempts have usually substituted norm-referenced scores for cut-off scores while continuing to use behavioral objectives. Consequently, only half a solution is provided. DRT and its accompanying instructional methodology offers constructive alternatives to both CRT's use of behavioral objectives and performance standards. Unfortunately, domain referenced measures of performance are few in number and are poorly understood. Communicating scores in terms of skill domains to the general public has proven difficult.

A possible solution to current educational measurement problems is the combined use of certain CRT characteristics, particularly those which DRT has improved, and NRT characteristics. The grouping of observable skills into concrete domains and the substitution of normative references for mastery scores would help provide the information needed for day-to-day instruction. Teachers not only would have the needed descriptions of test content with which to plan and monitor instructional programs, but also would be able to use the more familiar normative indexes to communicate these results. This particular combination of measurement characteristics also would provide educators with the information they need for prediction and selection purposes. Furthermore, by recognizing the fact that all tests and items should produce variance in performance, proven classical measurement approaches to determining reliability and validity could be used.

To further enhance the instructional usefulness of such a measurement device, scoring should include process, as well as the usual product scores. <u>Product scores</u> are quantitative indexes of performance and usually indicate whether or not, or to what extent, a task was accomplished. In educational measurement, product scores usually represent the number or percentage of items that were passed or failed. However, in addition to number, product scores can also represent distance, weight or volume.

<u>Process scores</u> are qualitative indexes of performance and usually provide descriptive information. Specifically, process scores could indicate what particular behaviors were or were not used in performing a certain task. For example, did the student

who failed to play a certain note on the violin do so because of incorrect fingering, or incorrect bowing, or because a combination of both behaviors prevented it. In this situation, a product score only would tell the teacher that the child was unable to perform the task. The teacher would not know why the failure occurred or how to remedy it. However, a scoring system which combined both product and process scores would have indicated that a failure had occurred, and what behaviors important to its accomplishment were or were not present.

To support this theoretical perspective of educational measurement, a concrete application of the arguments was attempted. Specifically, a test was constructed which would combine the various characteristics described here. The area of early childhood education was chosen to examine the feasibility of such an exercise.

<u>Assessing Motor Development in Early Childhood</u>

Many evaluations of various early intervention efforts have taken place over the past 15 to 20 years. However, few have provided information leading to a better understanding of the effects of intervention efforts on the growth and development of the preschool child. Some recent reviews of evaluations of early intervention programs (Ryan, 1972; Bronfenbrenner, 1975) have shown that the information provided by these evaluations have been restricted to the cognitive area and to only those skills which are measured by standardized intelligence or achievement tests. As a result, many important aspects of development are ignored, other than those provided by these cognitive measures. Furthermore, intervention efforts, designed to raise scores on such instruments, may be made while inhibiting other aspects of development. Evidence is needed to answer these questions and new measurement devices are needed to provide this evidence.

Within early childhood education, the area of motor development seems particularly in need of new assessment procedures. Halverson (1971) pointed out that until recently most information concerning motor development available to teachers was restricted to summaries of reports (e.g., Gesell & Thompson, 1938; Guttridge, 1939; Halverson, 1931; McGraw, 1935; Shirley, 1931; Wellman, 1937) completed in the 1930's. While these early studies did provide important beginnings at looking at how children develop physically, their findings were limited by poor methodology. For example, often the number of children comprising the subject pool was small and represented a limited segment of the population. Furthermore, the descriptions of the tasks the children were asked to perform were either vaguely written or nonexistent.

Early motor scales often reflected the faults of the motor development research. Many times, specific administrative directions or behavioral descriptions were not included. This aspect of the early motor scales reduced their validity since no one reading the manuals could clearly know what specific behaviors were being measured. The vague description of behaviors also eliminated the possibility of comparing motor items across scales, limiting the progress that such comparisons might provide to a general knowledge base. Furthermore, any research using such scales could provide only questionable findings.

Literature review. While the number of instruments have increased significantly since the 1930's, educators are still faced with basically the same problems that plagued their earlier counterparts. That is, few instruments are available which can provide reliable and valid information concerning an individual's motor performance which could be used for everyday instructional purposes.

Several sources in the literature were reviewed to determine the number and kinds of instruments available for the assessment of preschool motor development. Tests and Measurement in Child Development: Handbook I (Johnson & Bommarito, 1974) was used as was the two-volumed Handbook II

(Johnson, 1976). The later references included a search of 148 professional journals since 1966. In addition, the comprehensive bibliography from Cross and Goins' (1977) recent efforts was used. This bibliography describes all instruments used by programs in the Handicapped Childrens Early Education Program (HCEEP). Furthermore, Herkowitz's (1978) current review of instruments used to assess motor development was found helpful.

As a result of this effort, 42 instruments were identified which yield an index of motor functioning for some part of the three through five age range. These 42 instruments were categorized into three major categories: (1) standardized instruments, (2) nonstandardized instruments, and (3) experimental instruments. These major categories were subdivided into norm-referenced or criterion-referenced categories.

To be classified as a standardized instrument, clear administrative procedures had to be a part of the instrument. Similarly, scoring principles had to be unambiguously spelled out, and reliability and validity data had to be reported. In addition, norm-referenced tests needed to report the characteristics of the norming sample. If an instrument failed to meet any of these criteria, and was currently being used, it was classified as a nonstandardized instrument. The term experimental was applied to those instruments which are currently being developed, and are not being used in applied situations. A listing of all these preschool motor assessment instruments can be seen in the Preschool Motor Assessment Bibliography (Appendix A).

Examining the group of standardized instruments; several limitations can be seen. First, since few "normal" motor behaviors are included (e.g., walking, kicking, climbing, catching), the use of these instruments is restricted with nonhandicapped or mildly handicapped children. Second, only one standardized instrument, The Modified Lincoln-Oseretsky Motor Development Scale, was solely concerned with measuring motor development. While the other scales incorporated some form of motor assessment, they also evaluated other aspects of growth such as cognitive or social development. As a result, the number of motor behaviors sampled is often limited.

The educational limitations of the standardized instruments should not be ignored. Specifically, the scoring systems of these instruments primarily provide product rather than process indexes of performance. In addition, test content is not organized into, or even related to concrete domains of behavior. Both of these last two characteristics limit the usefulness of these instruments in planning daily instructional programs.

Confusion in the use of the term criterion-referenced testing (CRT) and norm-referenced testing (NRT) is widespread within the group of nonstandardized tests. Some of the authors of these tests have identified the test as criterion-referenced on the basis of the alignment of items to specific behavioral objectives (e.g., Brigance, 1978; Sanford, 1975; Umansky, 1974; Wendt, Schramm & Schmaltz, 1975). Still others have combined the CRT characteristic of well-defined skill behaviors with normative guidelines as "criteria"; (e.g., Dickerson, Evanson, Spurlock, 1975; Donahue, Montgomery, Keiser, Roecher, & Smith, 1975; Sanford, 1975; Sharp, 1975).

The misinterpretation in the meaning of CRT and NRT by the authors of the group of nonstandardized instruments has produced serious technical flaws. For example, many of the nonstandardized criterion-referenced tests neglect the specification of clear administrative directions. Even equipment specifications at times is left to whatever the examiner can find readily available. Furthermore, passing criteria are often poorly delineated. The nonstandardized norm-referenced counterparts do not offer many better options. Often these instruments are identified as screening devices with items collected from other standardized instruments (e.g., Anderson, Miles & Matheny, 1963; Meisels & Wiske, 1976). Unfortunately, few authors of these scales have chosen to restandardize their particular collection of items with different norming samples.

While the number of <u>experimental instruments</u> is very small, they do provide hope for the motor assessment area. The instruments are solely concerned with motor skills and include behaviors that permit the assessment of nonhandicapped children. Furthermore, the descriptions of the behaviors are clearly described. Necessary equipment and testing conditions are unambiguously specified and attempts have been, or are being made, to collect reliability and validity data.

Perhaps the most promising characteristic of these tests is the focus on educational usefulness. This is particularly true of the <u>DeOreo Fundamental Motor Skill Inventory</u> (DFMSI) as described by Herkowitz (1978). DeOreo designed this instrument to provide process statements as well as product statements (Halverson, 1971). For example, while indicating if a child could broad jump a specified distance (product statement), the instrument also describes how that skill was performed (process statement). Specifically, the process statement might indicate whether or not the child integrated the use of arms with the legs to add momentum to the jump.

Halverson (1971) and others (e.g., Hellebrandt, Rarich, Glassow & Carns, 1961) have argued that product statements (e.g., measures of distance, height, and number) provide only partial information. Furthermore, Halverson believed that the predominat use of product statements over the years has resulted in little being learned about how a young child progresses in the beginning stages of movement control. Halverson suggested that this problem could be remedied when motor assessment instruments were designed to accurately and fully describe movement.

In summary. The review of preschool motor assessment literature has indicated a need for new assessment devices within early childhood education. Early attempts to measure movement produced instruments of little use to research or instructional efforts. The behavioral descriptions included in these early instruments were often obscure and standardization was inadequate. While later instruments have improved standardization characteristics, they have several shortcomings. Specifically, few instruments are solely concerned with the assessment of motor behaviors. Furthermore, many of the more recent motor assessment instruments have not included behaviors which would permit the study and assessment of normal (i.e., nonhandicapped) development. Most importantly, preschool motor tests have not organized content and scoring processes to help educators plan instructional programs so that experiences designed to enhance motor development can be provided.

The Proposal

In response to the conclusion drawn regarding the field of educational measurement and preschool motor assessment, a study was designed. The feasibility was examined of designing a scale with both NRT and CRT characteristics which would reliably measure motor skills of children, ranging in age from three to six years.

To meet the purpose of the study, the following tasks were identified:

(1) selection of appropriate motor skills,

(2) organization of motor skills into behavioral domains,

(3) the design of administrative and scoring procedures,

(4) the administration of the experimental test to a sample of preschool age children.

The selection of appropriate motor skills included the use of certain criteria: (1) only discrete observable behaviors were included; (2) these behaviors must occur in a relatively sequential manner over the three to six year age range; (3) both gross and fine motor skills must be represented; and (4) behaviors were selected which would allow the assessment and study of both normal and deviant patterns of development.

The organization of motor skills into domains served several purposes. First, characteristics common to specific skills would be obvious and therefore assist measurement/observation processes. This organization should also provide a clear description of content and thus help the teacher who needs to design or redesign instructional programs in response to test data.

Administrative and scoring procedures were explicitly written to produce consistent, useful results. Both product and process scores were included.

The administration of the test to a sample of preschool-age children was conducted to determine empirically (1) what, if any, developmental sequences existed; (2) if any sex differences existed for each behavioral skill; (3) if the test could produce reliable, i.e., consistent, information; (4) if separate relatively independent skill domains could be verified.

METHODS

Subjects

Subjects were 83 children from the central Iowa area. They ranged in age from 36 to 70 months. There were 35 males and 48 females: 23 male and nine female five year olds; 10 male and 25 female four year olds; two male and 14 female three year olds. Socio-economic backgrounds ranged from children whose families qualified to receive federal or state economic support to children from professional families in a university community.

Testing Materials

Selection of motor skills. Using the criteria specified earlier for the selection of motor skills, 26 directly observable motor behaviors were selected for the motor test. These skills were ones which develop in a sequential manner over the three to six age range and represent 10 commonly recognized functional areas of motor development. Each motor skill functional area, and domain is listed in Table 1. These motor skills comprised what was labeled as the Preschool Perceptual Motor Scale (PPMS).

The functional areas of copying, dressing, and writing comprised the fine motor activities, and all other functional areas represented gross motor activities. Six of the functional skill areas, i.e., throwing, catching, kicking, dressing, jumping, and climbing, measured primarily normal developmental

Table 1. Breakdown of motor domains into functional skill areas and separate motor behaviors.

I. Stationary Gross Motor Domain

A. Functional area: Balancing
1. Right foot lifted
2. Left Foot Lifted

B. Functional area: Throwing
1. Small ball throw
2. Large ball throw

C. Functional area: Kicking
1. Ground level kick
2. Drop kick

D. Functional area: Catching
1. Throw - catch
2. Bounce - catch

II. Manipulative Small Motor Domain

A. Functional area: Writing
1. "Name" writing

B. Functional area: Copying Designs
1. Copying a cross
2. Copying a circle
3. Copying a square
4. Copying a triangle
5. Copying a rectangle
6. Copying a rectangular design

C. Functional area: Dressing
1. Snapping
2. Zipping
3. Buttoning
4. Lacing
5. Tying

III. Mobile Gross Motor Domain

A. Functional area: Walking
1. Board walking

B. Functional area: Climbing
1. Climbing upstairs
2. Climbing downstairs

C. Functional area: Jumping
1. Both feet (Broad jump)
2. Left foot hop
3. Right foot hop

motor patterns (e.g., Cratty & Martin, 1956; DeOreo, 1976; Hellebrandt, Rarich, Glassow & Carns, 1961). The copying, writing, balancing and walking skill areas were chosen because of their association with studies of both exceptional and normal behavioral patterns (e.g., Bender, 1938; DeOreo, 1976; Frostig, 1961; Getman, 1952; Kephart, 1964; Loovis, Note 9).

Domain construction. Baker (1974), Hively (1974), and Millman (1973) have discussed various approaches to constructing domains. These discussions were considered in formulating guidelines for the construction of the motor domains of the PPMS. First, since motor skills were selected prior to the forming of domain boundaries, an inductive method was used to form the motor domains (Hively, 1974). After careful comparison, the 10 functional skill areas were organized into three domains (Table 1.).

To specify the parameters of each set of related behaviors, written domain descriptions were included in the PPMS (Appendix B). This descriptive statement was not only intended to clarify content, but to direct the attention of the test user to a particular set of behaviors. However, generalizability of content was also emphasized. This emphasis was given in writing domain descriptions so that the classroom teacher could easily choose similar behaviors not specifically included in the test. This characteristic of the PPMS was designed to avoid the necessity of teaching to test items when a student performed poorly and needed remedial work. To enhance this generalizability of domain content, concrete examples of behavior which were illustrative of domain limitations were included in the descriptions.

Since domains and their written descriptions have been compared to behavioral objectives (e.g., Baker, 1974), statements of intent were avoided. Baker (1974) preferred emphasizing descriptive characteristics and neglecting intent, since measures "...may also be written to measure...other than that encompassed in the goals of an instructional program" (p. 22). This preference was followed in the specifications of the three domains of the PPMS.

Administrative guidelines. To enhance the standardization of the PPMS, equipment specifications for each motor task were included in the test. In addition, written directions for administering items and collecting data were included to aid the examiner. Domain descriptions and limitations, as well as equipment specifications and directions, can be seen in a copy of the PPMS (Appendix B).

Procedures

Item administration. All 26 items of the PPMS were administered to the subjects individually by the same experimenter on two separate occasions one week apart. The tasks were administered in the preschool or day care center the children attended. The children were told that they were going to play some games. The experimenter explained that some of the games would be easy and some would be difficult, but they should try their best on all games. No other directions except those stipulated in the test itself were given (see Appendix B).

To maximize interest in the activities and promote the possibility of completing the full test without breaks, functional skill areas of the PPMS were administered in the following order; (1) writing, (2) copying, (3) balancing, (4) walking, (5) climbing, (6) throwing, (7) catching, (8) kicking, (9) jumping, (10) dressing. All children completed the tasks without breaks in approximately 20 minutes.

Normative data was gathered regarding quantitative aspects of the children's performance (i.e., product information). Depending on the particular task, time duration, distance, or number of tasks were recorded. Using Halverson's (1971) suggestions, data concerning qualitative aspects of performance were also collected. These process data were collected only when information in the literature indicated what aspects of performance to observe. The product and process data were combined to provide a score for each of the 26 tasks. These scores were then combined across functional areas and domains to provide (1) functional area scores, (2) domain scores, and (3) a total score for each child.

Design

To determine the stability of scores over a week's time, product-moment correlation coefficients by age groups (three, four, and five years of age) were computed on task scores, functional area scores, and domain scores. To determine internal consistency, task scores on the first testing occasion were used to compute Cronbach's coefficient alpha (Nunnally, 1967). Task scores on occasion one also were factor analyzed to further examine scale homogeneity.

Finally, to determine if age, sex or their interaction resulted in statistically significant differences in performance, task scores served as dependent variables in 26 separate regression analyses of variance (ANOVA). Age and sex of the child served as independent variables in these ANOVAs and were combined factorially in a 2 (sex) X 3 (age) design.

RESULTS

To examine the feasibility of constructing a test with both NRT and CRT characteristics, statistical and noninferential analyses were conducted. The nonstatistical analysis was composed of comparing normative data from the administration of the PPMS with comparable data from the literature. Simple descriptive statistics were computed for the qualitative and quantitative aspects of the motor achievements on the children in this study. This normative data was compared in a logical, rather than a formal statistical fashion with normative data from the literature. From these comparisons product and process scores were derived for the various motor achievements of the children. These scores then were used in the three statistical analyses.

The three statistical analyses examined (1) test reliability, (2) the inter-relationships of motor tasks, functional areas and domains, and (3) the effects of sex and age of subjects on task performance. The normative data from the PPMS will be reported first, followed by test reliability data, the inter-correlation of test components, and the analyses of variance results describing age and sex differences in task performance.

Normative Data

The normative data in the following sections describes the various quantitative and qualitative aspects of motor performance on the PPMS. It also specifies what, if any, developmental trends were found. For a detailed explanation of how this normative data from the literature and the PPMS were used to derive process and product scores, see the section on scoring in Appendix B.

Stationary gross motor domain. Under the functional area of balancing, time to the nearest tenth of a second was recorded for the two tasks of balancing on one foot. Data from Bayley's *Preschool Scales of Motor Development* (as cited in Nelson, Note 10) indicated preschool-age children gradually are able to increase their time for this task over the three to five year age range. To determine how the present study's results compared to Bayley's data, a distribution of time durations was plotted by age and testing occasion. The descriptive statistics of these distributions can be seen in Table 2.

The mean and median time durations for both balancing tasks increased across age groups. Similarly, variability in performance increased across age groups. While time durations were somewhat less for the left foot balance task than the right foot balance task, they were not appreciably so.

Whether or not assistance (e.g., touching the wall) was used while performing the task was also recorded. Percentage of children using assistance by age group and occasion can be seen in Table 3. The percentages of children using assistance varied

Table 2. Distributions of time durations (in seconds) for balancing tasks.

Group	Right Foot Balance Occasion 1	Right Foot Balance Occasion 2	Left Foot Balance Occasion 1	Left Foot Balance Occasion 2
5 year olds (n=32)				
Mean	12.80	16.53	14.15	17.89
Standard deviation	9.81	15.14	14.03	15.49
Median	11.30	11.60	9.80	15.30
Range	1.0 - 35.0	1.0 - 61.0	1.0 - 50.5	1.5 - 62.5
4 year olds (n=35)				
Mean	6.90	7.85	9.19	7.58
Standard deviation	6.92	11.20	8.87	6.28
Median	3.40	3.80	5.00	5.00
Range	1.0 - 28.5	1.0 - 59.0	0.5 - 31.8	1.5 - 30.0
3 year olds (n=16)				
Mean	2.63	2.75	2.14	3.11
Standard deviation	2.43	2.61	1.33	4.53
Median	1.50	1.90	2.00	2.00
Range	0.0 - 8.6	0.0 - 10.0	0.0 - 5.2	0.0 - 19.5

Table 3. Percentage of children needing assistance for balancing tasks.

	RIGHT FOOT BALANCE		LEFT FOOT BALANCE	
	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)	9	3	9	9
4 year olds (n=35)	26	6	23	6
3 year olds (n=16)	38	31	31	19

Table 4. Distributions of distances (in inches) for throwing tasks.

Group	Small Ball Throw Occasion 1	Small Ball Throw Occasion 2	Large Ball Throw Occasion 1	Large Ball Throw Occasion 2
5 year olds (n=32)				
Mean	146.97	128.19	105.34	109.13
Standard deviation	55.74	42.56	43.80	47.01
Median	145.00	122.50	93.00	96.50
Range	70 - 274	62 - 216	34 -213	40 - 230
4 year olds (n=35)				
Mean	123.83	111.49	93.11	94.23
Standard deviation	54.23	49.85	45.93	42.45
Median	123.00	91.50	89.50	89.50
Range	24 - 234	32 - 216	39 - 264	36 - 234
3 year olds (n=16)				
Mean	87.13	89.69	55.06	63.25
Standard deviation	36.53	34.83	12.96	28.92
Median	72.50	83.50	54.00	54.50
Range	13 - 147	36 - 144	31 - 72	12 - 117

across occasions. However, the percentages were consistently low for all three age groups.

Under the functional area of throwing, distance to the nearest inch was recorded on three trials for the small and large ball throws. The three measurements for each task were averaged and distributions were plotted by age and testing occasion. The descriptive statistics of these distributions can be seen in Table 4.

There were increases in mean and median performance across age groups for both throwing tasks. Between the three-year-olds and the four-year-olds, inter-individual variability of performance increased an average of approximately 16 inches. However, a much smaller average increase of variability (two inches) occurred between the four and five-year-old groups. The same general pattern in variability can be seen in the large ball throw. All three age groups threw the small ball further than the large ball. The three-year-old group showed difference in variability between the two throwing tasks; however, no such difference in variability was found in the four and five-year-old groups.

Qualitative aspects of throwing were also recorded for the throwing tasks. Wild (1938) indicated that throwing a small ball follows a particular stage sequence of development. Initially a child stands in a stationary position and throws the ball underhanded with little shift in body weight. Gradually, the child's throwing skills progress so that other parts of the body are incorporated into the throw. First, the child rotates the body and steps with a foot while throwing. However, coordination of the body with the throw is not always initially efficient. Often the child steps with the foot which is on the side of the body from which the ball is thrown. Gradually, the child progresses so that the foot used in stepping is the one opposite the side of the body from which the ball is thrown. Eventually, a pronounced weight shift is seen as the child steps as s/he releases the ball.

Similar developmental patterns occur in learning to throw the large ball. First the child stands in a stationary position and throws the ball underhanded. Gradually, the child progresses so that s/he steps with the throw. Finally, the ball is thrown overhanded, coordinating the body by stepping into the throw.

Percentages of children using the three observable behaviors for the two throwing tasks on two of three trials were computed. These data can be seen in Table 5. Over 60 percent of all three age groups threw the small ball overhanded on both occasions. However, a smaller percentage

Table 5. Percentage of children using three behaviors with throwing tasks.

	SMALL BALL THROW			LARGE BALL THROW		
Behavior	OHT	SWT	SWOS	OHT	SWT	SWOS
Occasion 1						
5 year olds (n=32)	78	56	47	75	44	44
4 year olds (n=35)	63	26	6	57	31	31
3 year olds (n=16)	81	31	19	75	19	19
Occasion 2						
5 year olds (n=32)	81	47	34	81	28	31
4 year olds (n=35)	69	26	11	74	29	29
3 year olds (n=16)	75	19	0	75	0	0

OHT = Overhand throw.
SWT = Steps with throw.
SWOS = Steps with opposite side while throwing.

stepped with the throw, and even less used the foot opposite the side of the body from which the ball was thrown.

The age groups performed in a similar manner when throwing the large ball. The percentages of children in each group stepping with the throw were less than the percentage of children using an overhand throw. But the percentage of each group that stepped with the throw was the same as the percentage of children in each group that stepped with the foot opposite the side of the body from which the ball was thrown.

Under the functional area of kicking, qualitative and quantitative data were collected. Specifically, for both tasks data were recorded concerning whether or not the ball was kicked in a forward direction on three trials (quantitative data). In the drop kick task, each child was observed to see if s/he coordinated the release of the ball with the kicking of the foot (qualitative data). For example, some children waited to kick the ball until it reached the floor. Others stooped down and held the ball with their hands while kicking it. Table 6 shows the percentage of children, by age group, who successfully completed these tasks.

A high percentage of all three age groups successfully completed the ground level kick; in fact, 100 percent of both the five year old and four year old groups completed

Table 6. Percentage of children successfully completing kicking tasks.

	GROUND LEVEL KICK		DROP KICK			
	Kicked in Forward Direction		Kicked in Forward Direction		Coordinated Kick	
	Occasion 1	Occasion 2	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)	100	100	40	31	47	41
4 year olds (n=35)	100	100	25	40	31	51
3 year olds (n=16)	87	93	13	13	13	13

the tasks on both occasions. Only two three-year-olds were unable to successfully complete the tasks on the first testing occasion, and only one failed on the second occasion.

In each age group the percentage of children who were able to drop kick the ball in a forward direction was low; in fact, the highest figure on both occasions was 40 percent. Percentages varied across occasions for the five and four year olds, but remained the same for the three year olds.

For the four and five year olds, percentages of children coordinating the release of the ball with the kick of the foot were a little higher than the percentages of children successfully kicking the dropped ball. That is, these children did not try to kick the ball too soon nor did they wait too long. Yet, they failed to kick the ball. However, the same percentage of children who kicked the ball in a forward direction also coordinated the release of the ball with the kick on both occasions.

Under the functional area of catching, again qualitative and quantitative data were recorded. Specifically, whether or not the ball was caught on three trials was recorded for both tasks (quantitative data). The arm position used by the child for the tossed-ball catch was also recorded on three trials (qualitative data).

Cratty (1970) has described studies which have found three levels of development for catching behavior that occur in the preschool-age range. The first level is where the child's arms are held in a circular "cradle-like" fashion. At the second level, the child extends both arms with stiff elbows in front of the body while catching. Finally, the child advances to the level where the arms are relaxed and to the sides of the body, ready to respond to the object being thrown. Usually, accuracy of catching increases with increases in developmental level.

Percentage of children successfully catching the ball on two of three trials on both tasks, and the percentage of children in each age group using the different arm positions for the toss-catch were figured.

Table 7 shows the percentage of children by age group who successfully caught a tossed and a bounced ball. When the ball was tossed to the child, a high percentage of all age groups caught it. When the ball was bounced, somewhat different results occurred. A high percentage of children caught the bounced ball in all but the three-year-old group.

Table 8 shows the percentage of children that used the

Table 7. Percentage of children catching a tossed and a bounced ball.

	TOSS - CATCH		BOUNCE - CATCH	
	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)	100	100	94	84
4 year olds (n=35)	91	97	77	77
3 year olds (n=16)	75	63	38	44

Table 8. Percentage of children using different arm positions to catch a tossed ball.

	Cradle Position		Stiff Arm Position		Relaxed Arm Position	
	Occasion 1	Occasion 2	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)	15.6	15.6	40.6	43.75	43.75	40.6
4 year olds (n=35)	20.0	31.4	54.3	45.7	25.7	22.9
3 year olds (n=16)	31.2	6.3	62.5	87.4	6.3	6.3

different arm positions to catch the tossed ball. For the five-year-old group, a larger percentage of children used the stiff-arm position than used the cradle-arm position. However, there is little difference between percentages of five year olds using the relaxed arm position and the stiff-arm position. For the four year olds, a greater percentage used the stiff-arm position than used either of the other two arm positions. The percentage of four-year-old children using the cradle arm position was quite similar to the percentage using the relaxed arm position. The percentages of the four and five year olds did not vary much between testing occasions.

The three year olds used the relaxed arm position less than the other two arm positions on both occasions. The same percent of three year olds used the cradle position, but only on the second testing occasion. A high percent (31.2%) of this group used the cradle position on the first testing occasion. Finally, the three year olds used the stiff arm position more than the other two arm positions.

Comparing across age groups, more five year olds than four year olds used the relaxed arm position. Likewise, a greater percentage of four year olds than three year olds used the relaxed-arm position. The opposite trend can be seen when considering the use of the stiff-arm position.

More three year olds than four year olds used the stiff-arm position and a greater percentage of four year olds than five year olds used this position. A similar pattern occurred for the cradle-arm position; however, the level of percentages were somewhat lower and more variable over testing occasions.

Manipulative small motor domain. Under the functional area of writing only qualitative data were gathered. Hildreth's (1936) study of "name" writing of preschool-age children provided some guidance as to what particular developmental aspects of writing were important to observe. Specifically, the children's name writing was examined to determine if only scribbles were present, or if letters were mixed with scribbles. If only scribbles were used, they were further examined to see if they were random and circular, or to what extent horizontal movement and vertical strokes were used. The use of capital letters and lower case letters also was recorded. In addition, the presence of letter reversals and misspellings was recorded. Table 9 shows the percentage of children exhibiting the various "name" writing characteristics.

The same percentage of each age group that scribbled while writing their name, did not scribble in a random, circular fashion. That is, while scribbling was present among the name writing tasks, all showed some degree of horizontal movement or vertical strokes. More three year olds than four

Table 9. Percentage of children using name writing characteristics.

	Scribbles vs. Letters				Capital vs. Lower Case			Reversals		Spelling
Behavior	SO	AS	MSL	AL	ACL	LCLO	MCLC	LR	WR	CS
Occasion One										
5 year olds (n=32)	0	0	6	94	34	3	63	22	3	84
4 year olds (n=35)	11	11	20	69	49	3	13	20	6	54
3 year olds (n=16)	56	56	13	31	31	6	13	25	13	19
Occasion Two										
5 year olds (n=32)	0	0	6	94	34	3	63	22	3	84
4 year olds (n=35)	11	11	20	69	49	3	13	20	6	54
3 year olds (n=16)	56	56	13	31	31	6	13	25	13	19

SO = Scribbles Only
AS = Aimless or Circular Scribbling
MSL = Mixture of Scribbles and Letters
AL = All Letters
ACL = All Capital Letters
LCLO = Lower Case Letters Only
MCLC = Mixture of Capital and Lower Case
LR = Letter Reversals
WR = Word Reversals
CS = Correct Spellings

year olds, and more four year olds than five year olds used only scribbles. Conversely, a greater percentage of five year olds than four year olds, and more four year olds than three year olds used only letters.

The same pattern of intergroup percentages that appeared for scribbling was seen in the use of letters. Specifically, a larger proportion of the older groups used a mixture of capital and lower case letters. In comparison, more of the younger groups used solely capitals or lower case letters. Similarly, the two older groups reversed words less often. However, letter reversals remained very similar across the three age groups. In examining spellings, once again the older the group of children, the larger the percentage of correct spellings was found. Identically the same writing characteristics were found for each age group on the second testing occasion.

Under the functional area of copying designs, several performance characteristics were observed and recorded. Scoring schemes, such as those used with the Bender Motor Gestalt Test, were used as a reference to determine what aspects of copying should be recorded. Specifically, characteristics which define each geometric design were observed. That is, the circle drawings were examined to see if they were circular rather than oval and to see if they were open or closed. Drawings of the cross were inspected to see if the lines intersected, while figures with corners (i.e., squares, triangles, and rectangles) were examined to see if corners were rounded or not. In addition, the length of the sides of the squares, triangles, and rectangles was examined. Finally, the general orientation of the copied figures was observed. Table 10 shows the percentage of each age group displaying these characteristics in the various drawings.

The older the child, the more likely the drawings approximated the stimuli on both occasions. Specifically, a larger percentage of five year olds than four year olds, and more four year olds than three year olds, oriented their drawings as shown in the examples. Likewise, the older the group the more likely the defining characteristics of each figure were included in the drawings. The copying of the circle was the only task which did not produce similar results. Here, fewer five year olds than four year olds drew circular rather than oval shapes, and fewer closed the drawings. However, more four year olds than three year olds included these defining characteristics.

Approximately the same percentage of five year olds and four year olds included all the defining features when copying the cross and the circle on both occasions. All other copying tasks had lower percentages of four and five year olds including

Table 10. Percentage of children using defining features when copying geometric designs.

	Geometric Designs																					
	Cross			Circle			Square			Triangle			Rectangle				Rectangle-Diagonal					
Characteristic	LI	DO	BOTH	CL	ACC	BOTH	CO	Si	BOTH	CO	DO	BOTH	CO	Si	DO	ALL	CO	Si	DO	LP	LO	ALL
Occasion One																						
5 year olds (n=32)	100	84	84	84	84	84	100	69	69	78	63	63	94	81	81	69	97	53	81	94	59	31
4 year olds (n=35)	94	91	91	91	91	91	83	46	46	37	29	29	74	51	63	40	57	31	37	83	23	14
3 year olds (n=16)	75	75	69	81	81	81	13	13	13	6	6	6	19	13	6	6	13	13	6	19	6	6
Occasion Two																						
5 year olds (n=32)	100	91	94	81	81	81	100	66	66	91	66	72	100	75	97	72	91	66	81	100	63	50
4 year olds (n=35)	94	94	94	97	97	97	86	60	60	43	37	37	80	60	54	37	63	37	54	83	23	17
3 year olds (n=16)	75	63	63	81	81	81	25	6	6	6	6	6	13	19	6	6	19	13	13	50	0	0

LI = Lines Intersect
DO = Drawing Orientation
CL = Closure of Circle
ACC = Accuracy of Drawing
CO = Corners of Drawing

Si = Sides of Drawing
LP = Line Present
LO = Line Orientation
BOTH = Both Characteristics Included
ALL = All Characteristics Included

all defining characteristics of the designs. The task of copying the rectangle with a diagonal line had the lowest percentages of the two older age groups including all defining features.

The intertask percentages of the three year olds for including all defining characteristics of a design displayed a different pattern. On both occasions, more three year olds included both defining features in the drawing of the circle than in the drawing of the cross. All other copying tasks had approximately the same percentages.

Since information concerning dressing is practically nonexistent in the literature, only quantitative aspects of performance were recorded. Specifically, whether or not the child latched the zipper and zipped it up and down was recorded. The number of buttons buttoned and unbuttoned, and the number of snaps snapped and unsnapped also was recorded. Similarly, the number of eyelets laced and the number of bows used in tying were recorded. On the tying task, whether or not the child completed a simple overhand knot and whether or not bows were used was observed. Percentage of children in each age group completing these various dressing tasks can be seen in Table 11.

Intergroup percentages indicated that the older the child, the more likely the particular dressing skill would be completed.

Table 11. Percentage of children completing dressing tasks.

Task	Zipping		Buttoning				Snapping				Lacing		Tying		
	LZ	ZUD	BA	B≤½	UBA	UB≤½	SA	S≤½	USA	US≤½	LA	L≤½	TK	T1B	T2B
Occasion One															
5 year olds (n=32)	78	100	100	0	100	0	100	0	100	0	66	31	84	66	47
4 year olds (n=35)	60	100	83	17	94	3	97	3	100	0	60	40	60	34	17
3 year olds (n-16)	25	100	50	50	63	37	63	37	81	29	13	87	19	0	0
Occasion Two															
5 year olds (n=32)	91	100	100	0	100	0	100	0	100	0	81	19	88	75	59
4 year olds (n=35)	69	100	94	6	94	3	97	3	100	0	50	37	74	46	23
3 year olds (n=16)	25	100	63	50	75	19	69	31	100	0	6	93	19	0	0

Dressing Task

LZ = Latches Zipper
ZUD = Zips Up and Down
BA = Buttons All Buttons
B≤½ = Buttons ½ or Less of Buttons
UBA = Unbuttons All
UB≤½ = Unbuttons ½ or Less of Buttons
SA = Snaps All Snaps
S≤½ = Snaps ½ or Less of Snaps
USA = Unsnaps All Snaps
US≤½ = Unsnaps ½ or Less of Snaps
LA = Laces All Eyelets
L≤½ = Laces ½ or Less of Eyelets
TK = Ties Knot
T1B = Ties One Bow
T2B = Ties Two Bows

More five year olds than four year olds, and more four year olds than three year olds latched the zipper, buttoned and unbuttoned all buttons, snapped and unsnapped all snaps, laced all eyelets, tied knots, and tied two bows. This occurred on both testing occasions.

In the five-year-old group, percentages among tasks for completing dressing tasks were highest for zipping, buttoning and unbuttoning all buttons, snapping and unsnapping all snaps. Percentages were lowest for lacing all eyelets, and tying two bows. For the four year olds, percentages were highest for the four snapping and buttoning tasks. However, fewer four year olds than five year olds latched the zipper, laced all eyelets and tied a knot. The lowest percentage in the four-year-old group was for tying two bows (17%). The intertask percentages for completing dressing tasks in the three-year-old group were lower than in the four-year-old group, but exhibited the same relationships between tasks.

Mobile gross motor domain. Only quantitative data were recorded under the functional area of walking. In particular, whether or not the child walked the full length of the board was recorded on two separate trials. In addition, the time (in seconds) required to complete the walk was recorded on the second trial.

On the first testing occasion, 97 percent of the five year olds walked the full distance of the beam. A smaller percent (91%) of four year olds and three year olds (63%) completed the task successfully. On the second testing occasion, again 97 percent of five year olds walked the full length of the board, while 94 percent of the four year olds and 69 percent of the three year olds completed the task. A distribution of time in seconds to complete the walking task was plotted by age and testing occasion. The descriptive statistics of this distribution can be seen in Table 12.

Median time durations on both occasions increased sequentially across age groups with the five year olds exhibiting the shortest durations and three year olds showing the longest. Mean time durations were somewhat different. While group means were similar to median times in that they were lower for four year olds than five year olds no such pattern was seen in comparing four year olds and three year olds. In fact, approximately the same mean durations were found for both testing occasions. The group standard deviations and range statistics showed variability in time durations increasing across age groups.

Table 12. Distributions of time durations (in seconds) for board walking.

Group	Testing Occasion: Occasion One	Testing Occasion: Occasion Two
5 year olds (n=32)		
Mean	2.49	2.51
Standard deviation	2.24	1.55
Median	1.95	1.95
Range	1.0 - 8.0	1.1 - 6.0
4 year olds (n=35)		
Mean	4.39	4.56
Standard deviation	3.46	4.51
Median	2.80	3.30
Range	1.5 - 14.5	1.2 - 21.5
3 year olds (n=16)		
Mean	6.24	4.38
Standard deviation	6.01	5.00
Median	4.10	5.60
Range	2.6 - 16.0	1.0 - 13.3

Under the functional area of climbing, only qualitative data were recorded. Using data summarized by Espenschade and Eckert (1967) and Cratty (1970), two major aspects of stair climbing were recorded. The first characteristic focused on the use of support (e.g., handrail) while climbing up and down the stairs. The other aspect of stair climbing which was observed concerned the stepping patterns exhibited by the children. In particular, whether a mark time pattern or an alternating foot pattern was used was recorded.

The literature reviews by Cratty (1970) and Espenschade and Eckert (1967) indicated that stair climbing follows a sequential stage of development. Children initially require some sort of assistance or support when climbing up and down the stairs. Furthermore, climbing upstairs usually precedes climbing downstairs, and a mark time stepping pattern comes before the alternating foot pattern. Table 13 shows by age group the percentage of children in this study who accomplished the various stair climbing achievements.

Percentages among tasks generally showed that more children in all age groups used support when climbing downstairs than when climbing upstairs. Likewise, more children in each age

Table 13. Percentage of children accomplishing stair climbing achievements.

	Climbing Upstairs			
	With Assistance		Mark Time Pattern	
	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)	25	9	3	9
4 year olds (n=35)	23	23	6	9
3 year olds (n=16)	38	19	44	38

* *

	Climbing Downstairs			
	With Assistance		Mark Time Pattern	
	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)	28	13	19	13
4 year olds (n=35)	37	31	43	43
3 year olds (n=16)	44	38	69	81

group used the mark-time stepping pattern when climbing downstairs than when climbing upstairs. Comparing percentages across age groups, more three year olds than four year olds climbed both up and down the stairs with assistance and used a mark-time stepping pattern. However, this was not true when climbing upstairs. Here percentages between the two youngest groups were approximately the same.

Under the functional area of jumping, both qualitative and quantitative data were recorded. In particular, whether or not both feet left the floor at the same time was recorded for the broad jumping task (qualitative data). Additionally, the distances of the broad jump and the right and left foot hops were recorded (quantitative data). Statistics describing the various distributions of distances for each jumping task can be seen in Table 14.

Mean and median statistics for the three age groups showed five year olds jumped the furthest distances, and three year olds jumped the shortest distances, on all tasks on both occasions. Five year olds were most variable in their performance and three year olds were least variable.

Reliability

Three kinds of reliability information were examined to determine if the test scores of the PPMS produced consistent information. First, inter-rater reliabilities of scoring

Table 14. Distribution of distances (in inches) for jumping tasks.

	Broad Jump		Right Foot Hop		Left Foot Hop	
Group	Occasion 1	Occasion 2	Occasion 1	Occasion 2	Occasion 1	Occasion 2
5 year olds (n=32)						
Mean	33.81	31.84	162.34	165.56	121.97	144.34
Standard deviation	10.32	9.37	135.57	120.75	94.11	103.16
Median	35.50	34.50	144.00	144.00	128.50	145.00
Range	6 - 48	16 - 48	6 - 612	8 - 540	7 - 354	0 - 330
4 year olds (n=35)						
Mean	26.51	25.37	82.11	76.11	72.23	81.69
Standard deviation	9.97	7.29	73.05	70.80	73.61	73.31
Median	26.50	24.00	63.25	53.50	63.00	63.00
Range	10 - 44	13 - 42	0 - 288	0 - 264	0 - 327	0 - 267
3 year olds (n=16)						
Mean	15.25	16.75	4.38	10.63	6.00	10.13
Standard deviation	9.97	4.30	8.80	11.42	10.19	12.70
Median	16.50	17.50	1.00	9.33	5.00	7.00
Range	0 - 35	10 - 27	0 - 33	0 - 36	0 - 34	0 - 38

the copying tasks were collected. Second, test-retest reliabilities for all tasks over a one week period were analyzed. In addition, internal consistency estimates were computed using the 26 task scores. The reasons for each analysis and the results that were produced are detailed throughout the following section.

Inter-rater reliability. To determine the extent that the scoring criteria of the copying task were clearly specified and provided for consistent results, the scoring of two independent raters was analyzed (see Appendix B for scoring criteria). Product moment correlation coefficients were computed on the scoring of the raters for each of the six drawings on both testing occasions. Results revealed a high degree of agreement. No coefficient was below .87 and all were significant at the .01 level. For the drawings produced on the first testing occasion, coefficients of .96, .87, .97, .94, .95, and .93 were obtained for the scoring of the cross, circle, square, triangle, rectangle, and rectangular design drawings, respectively. Coefficients for the scoring of the second occasion drawings were .94, .90, .93, .88, .91, and .88 for the cross, circle, square, triangle, rectangle, and rectangular designs, respectively.

Internal consistency. Homogeniety was investigated to determine the degree to which the tasks on the PPMS interrelated and measured a single characteristic. In examining the content of the PPMS and determining the consistency of performance over all tasks included in the tests, two measures of homogeneity were used; (1) coefficient alpha, and (2) factor analysis. Using coefficient alpha, internal consistency estimates were computed for each age group on all task scores for the first testing occasion. Indexes of .76, .71, and .83 resulted for the five, four, and three year old groups, respectively. All were significant at the .05 level.

These reliability figures indicated a moderately high level of consistency of performance across tasks for each of the three age groups. While all three coefficients were approximately the same, the three year olds performed most consistently across items and the four year olds performed least consistently. These differences in performance between age groups across tasks may reflect developmental variability. The normative data from the PPMS indicated that in comparison to the three year olds, the five and four year olds possessed more of the skills necessary to successfully perform the various motor tasks. However, the two older groups probably have not fully refined these skills, and therefore perform somewhat inconsistently. On the other hand, the three year olds are probably at a less mature level of development and lack a majority of the necessary skills, therefore, they perform not only lower than the five and four year olds, but they perform consistently lower.

The second approach used to examine the homogeneity of the tasks was factor analysis. This technique was used to determine the minimum number of factors necessary to account for the interrelationship of the 26 task variables or to what extent a single characteristic of motor performance was measured. For example, if the tasks of each of the ten functional areas loaded on separate factors, this would be some empirical support for the organization of tasks. However, if the factor analysis indicated that one factor (or a small number of factors) was sufficient to account for the variation in performance on all tasks, the evidence would not support the grouping of tasks into ten functional areas or domains.

A principal components analysis was computed using the 26 task scores and the ages of the children. Ten factors were extracted through this method. In the unrotated matrix, the majority of the loadings on nine of the factors were below .30 and widely dispersed. However, over half the tasks loaded .60 or higher on one factor and nine other variables loaded between .30 and .59 on this factor. To help clarify the pattern of factor loadings, the 10 factors underwent an orthogonal rotation. This rotation resulted in a less clear situation. Thus, the results of the unrotated matrix are reported. Table 15 shows one factor accounted for approximately one half (35.9%) of the common variance extracted through the analysis.

Table 15. Factor matrix of task scores and age.

	1	2	3	4
Copy Cross	0.40	-0.29	-0.42	-0.28
Copy Circle	0.04	0.29	-0.07	-0.25
Copy Square	0.75	0.07	-0.09	0.14
Copy Triangle	0.73	0.20	-0.08	-0.15
Copy Rectangle	0.80	0.18	-0.05	-0.05
Copy Rectangular Design	0.78	0.23	-0.11	-0.12
Name Writing	0.75	-0.15	-0.10	0.01
Zipping	0.66	0.01	-0.29	0.07
Buttoning	0.64	-0.15	-0.38	0.03
Snapping	0.51	-0.51	-0.19	-0.28
Lacing	0.63	0.11	-0.17	0.33
Tying	0.69	0.00	-0.08	0.18
Board Walk	0.62	-0.36	0.11	0.06
Broad Jump	0.67	0.10	0.30	-0.02
Right Foot Hop	0.75	-0.01	0.17	-0.07
Left Foot Hop	0.70	-0.14	0.22	-0.06
Climbing Upstairs	0.41	-0.39	0.37	0.27
Climbing Downstairs	0.43	-0.41	0.33	0.45
Left Foot Balance	0.67	0.07	0.32	-0.21
Right Foot Balance	0.72	0.10	0.35	-0.18
Small Ball Throw	0.39	0.53	0.13	-0.01
Large Ball Throw	0.36	0.42	0.29	-0.19
Throw - Catch	0.20	0.24	-0.47	0.58
Bounce - Catch	0.49	0.17	-0.32	0.07
Ground Level Kick	-0.07	0.17	0.27	0.58
Drop Kick	0.34	0.05	-0.10	-0.20
Age	0.87	-0.04	0.03	-0.08
Eigenvalues	9.70	1.70	1.69	1.49
Percent of Variance	35.90	6.30	6.30	5.50

5	6	7	8	9	10	Communality h^2
-0.21	-0.09	0.13	0.32	0.16	-0.24	.754
0.59	-0.38	0.22	-0.11	-0.22	-0.41	.923
0.14	-0.17	0.09	-0.16	-0.07	0.19	.700
-0.15	0.08	-0.15	-0.25	0.04	0.02	.697
-0.05	-0.01	-0.07	-0.18	0.07	-0.08	.729
0.12	-0.17	-0.17	-0.25	-0.03	0.03	.810
0.04	-0.12	-0.10	-0.19	0.06	-0.03	.662
0.09	0.28	0.14	-0.17	0.09	-0.05	.670
0.14	-0.11	0.01	0.01	-0.14	0.23	.682
-0.11	-0.20	-0.15	0.08	-0.03	-0.24	.774
-0.25	0.18	0.35	-0.07	0.04	-0.24	.828
-0.13	0.32	0.27	-0.02	0.10	-0.11	.730
0.06	0.06	-0.38	0.19	-0.04	-0.21	.763
0.01	-0.02	-0.27	-0.05	-0.34	-0.07	.746
-0.10	0.30	-0.10	0.11	-0.09	-0.19	.763
0.16	0.35	-0.19	-0.07	-0.09	0.09	.767
0.00	-0.45	0.31	-0.01	-0.09	-0.06	.840
-0.14	-0.31	0.09	-0.07	0.09	0.21	.845
0.29	0.08	0.32	0.28	0.15	0.08	.901
0.26	0.16	0.30	0.19	0.13	0.12	.934
-0.46	-0.25	0.04	-0.05	-0.03	-0.03	.730
-0.41	-0.27	-0.13	0.33	0.11	-0.04	.807
0.10	-0.18	-0.07	0.29	0.05	-0.02	.789
0.11	0.00	-0.06	0.46	-0.34	0.31	.815
0.36	0.07	-0.28	0.17	0.19	-0.28	.799
0.23	-0.22	-0.24	-0.04	0.66	0.15	.787
-0.06	-0.03	-0.03	0.03	-0.07	0.13	.794
1.36	1.26	1.10	0.99	0.93	0.84	21.060
5.00	4.70	4.10	3.70	3.40	3.10	78.000

This indicated that the test was fairly homogeneous in content. The nine other factors combined to account for the other one half of the extracted variance, but no single factor accounted for more than 6.3 percent. Obviously, only one factor was necessary to account for most of the interrelations among tasks.

Test-retest reliability. To determine stability over time, product moment correlation coefficients were computed on each of four scores for the two separate testing occasions. The task scores of each functional area were summed to provide a functional area score, a domain score and a total score. The stability coefficients for these scores can be seen in Tables 16, 17, and 18 for domains one, two, and three, respectively.

The manipulative small motor domain produced the largest number (69%) of coefficients above .70. Over half of all the coefficients for this domain were .80 or above. In comparison, the stationary gross motor domain had only four coefficients above .70 with all other between -.28 and .68. The mobile gross motor domain was somewhat better with ten coefficients above .70 and the other coefficients ranging from -.06 to .67. Since the domain scores were composed of the various task scores their test-retest coefficients were generally higher than those for the separate task scores. Seven tasks showed perfect stability over the two testing occasions for at least one age group. They were (1) the

Table 16. Test-retest coefficients for scores from the Stationary Gross Motor Domain.

	5 Year Olds (n=32)	4 Year Olds (n=35)	3 Year Olds (n=16)
Functional Area: Balancing	.50	.62	-.28*
Left Foot Balance	.27	.61	-.06*
Right Foot Balance	.44	.46	-.19*
Functional Area: Throwing	.53	.72	.13*
Small Ball Throw	.52	.67	.07*
Large Ball Throw	.44	.68	-.03*
Functional Area: Kicking	.40	.29*	.30*
Ground Level Kick	1.00	1.00	.26*
Drop Kick	.40	.29*	.43*
Functional Area: Catching	.58	.64	.10*
Throw - Catch	1.00	-.06*	.45*
Bounce - Catch	-.10*	.56	.10*
Domain Score	.59	.83	.53

*Coefficients with an * have alpha levels above .05.

Table 17. Test-retest coefficients for scores from the manipulative small motor domain.

	5 Year Olds (n=32)	4 Year Olds (n=35)	3 Year Olds (n=16)
Functional Area: Writing	.98	.89	.83
"Name" writing	.98	.89	.83
Functional Area: Copying Designs	.32	.84	.86
Copy a cross	.53	.94	.85
Copy a circle	.23*	-.05*	1.00
Copy a square	.36	.64	.86
Copy a triangle	.29	.82	1.00
Copy a rectangle	.40	.49	.67
Copy a rectangle with a line	.70	.72	.74
Functional Area: Dressing	.83	.87	.81
Snapping	1.00	1.00	.59
Zipping	.61	.70	.67
Buttoning	1.00	.82	.81
Lacing	.72	1.00	.68
Tying	.80	.79	.59
Domain Score	.65	.82	.88

*Coefficients with an * have alpha levels above .05.

Table 18. Test-retest coefficients for scores from the mobile gross motor domain.

	5 Year Olds (n=32)	4 Year Olds (n=35)	3 Year Olds (n=16)
Functional Area: Walking	.77	.36	.74
Walking a board	.77	.36	.74
Functional Area: Climbing	.38	.52	.30*
Climbing upstairs	.36	.49	.25*
Climbing downstairs	.32*	.53	.46
Functional Area: Jumping	.77	.80	-.06*
Broad jump	.67	.47	.20*
Left foot hop	.73	.66	.00*
Right foot hop	.47	.52	.21*
Domain Score	.76	.77	.89

*Coefficients with an * have alpha levels above .05.

ground level kick, (2) the throw-catch task, (3) the copying of the circle, (4) the copying of the triangle, (5) the snapping tasks, (6) the buttoning tasks and (7) the lacing tasks.

Inter-correlation among Tasks, Functional Areas and Domains

To further determine if organizing tasks into functional areas and domains could be empirically justified, product moment correlation coefficients were computed among functional area scores, domain scores and total score. All of the scores commonly represent motor skills and should show some degree of relationship. However, to justify the organization of task scores into functional area or domain scores, relationships within domains should be higher than relationships across domains. Tables 19, 20, and 21 show the product moment correlation coefficients among functional area, domain, and total scores for the five year olds, the four year olds, and the three year olds, respectively.

Examining all three domains, no distinct differences can be found. The correlations of functional areas within domains were very similar to the correlations of functional areas across domains for all three age groups. The correlations among all functional areas ranged from very low to moderate (e.g., .01 to .54). In conclusion, this data did not empirically indicate that tasks could be organized into hierarchical categories.

Table 19. Correlations among functional area, domain, and total scores for four year olds (n=35).

	D:SGM	FA:B	FA:T	FA:K
Domain: Stationary Gross Motor (D:SGM)		.60	.79	.17*
Functional Area: Balancing (FA:B)			.10*	.09*
Functional Area: Throwing (FA:T)				.05*
Functional Area: Kicking (FA:K)				
Functional Area: Catching (FA:C)				
Domain: Manipulative Small Motor (D:MSM)				
Functional Area: Writing (FA:W)				
Functional Area: Copying Designs (FA:CD)				
Functional Area: Dressing (FA:D)				
Domain: Mobile Gross Motor (D:MSM)				
Functional Area: Walking (FA:WL)				
Functional Area: Climbing (FA:CL)				
Functional Area: Jumping (FA:J)				
Total Score[a]				

*Coefficients marked with * have alpha levels greater than .05.
[a]Total score is a summation of all functional area scores.

FA:C	D:MSM	FA:WR	FA:CD	FA:D	D:MGM	FA:WL	FA:CL	FA:J	TOTAL
.31*	.38	.09*	.45	.23*	.42	.23*	.10	.43	.66
-.12*	.29*	.20*	.27*	.25*	.33	.09*	.16*	.32*	.46
.11*	.27*	.04*	.37	.12*	.30*	.14*	.10*	.29*	.49
-.15*	.27*	.24*	.28*	.10*	.31*	.26*	.17*	.26*	.31*
	-.04*	-.25*	.04*	.00*	-.01*	.13*	-.24*	.06*	.07*
		.74	.93	.75	.64	.36	.24*	.62	.91
			.57	.42	.45	.41	.18*	.41	.61
				.51	.61	.31*	.19*	.41	.61
					.46	.22*	.18*	.45	.66
						.61	.42	.94	.82
							.07*	.54	.48
								.11*	.31
									.80

Table 20. Correlations among functional area, domain, and total scores for three year olds (n=16).

	D:SGM	FA:B	FA:T	FA:K
Domain: Stationary Gross Motor (D:SGM)		.50	.66	.34*
Functional Area: Balancing (FA:B)			-.02*	.25*
Functional Area: Throwing (FA:T)				-.22*
Functional Area: Kicking (FA:K)				
Functional Area: Catching (FA:C)				
Domain: Manipulative Small Motor (D:MSM)				
Functional Area: Writing (FA:W)				
Functional Area: Copying Designs (FA:CD)				
Functional Area: Dressing (FA:D)				
Domain: Mobile Gross Motor (D:MGM)				
Functional Area: Walking (FA:WL)				
Functional Area: Climbing (FA:CL)				
Functional Area: Jumping (FA:J)				
Total Score[a]				

*Coefficients marked with * have alpha levels greater than .05.
[a]Total score is a summation of all functional area scores.

FA:C	D:MSM	FA:WR	FA:CD	FA:D	D:MGM	FA:WL	FA:CL	FA:J	TOTAL
.33*	-.06*	-.01*	.08*	-.07*	-.06*	-.10*	-.11*	.08*	.31*
.11*	-.03*	.15*	.02*	-.25*	-.01*	-.09*	-.10*	.24*	.15*
-.19*	-.08*	-.12*	-.03*	-.05*	-.11*	-.13*	-.13*	.02*	.12*
.08*	.07*	.08*	.08*	-.01*	.07*	.18*	.18*	.29	.27*
	.16*	.01*	.16*	.17*	-.22*	-.09*	-.09*	-.34*	.15*
		.80	.87	.71	.35*	.24*	.40*	.09*	.89
			.53	.49	.65	.41*	.52	.53	.84
				.37*	.07*	.03*	.21*	-.15*	.71
					.29*	.26*	.33*	.02*	.62
						.71	.87	.64	.61
							.51	.23*	.40*
								.27*	.59
									.33*

Table 21. Correlations among functional area, domain, and total scores for five year olds (n=32).

	D:SGM	FA:B	FA:T	FA:K
Domain: Stationary Gross Motor (D:SGM)		.60	.83	.40
Functional Area: Balancing (FA:B)			.20*	.13*
Functional Area: Throwing (FA:T)				.14*
Functional Area: Kicking (FA:K)				
Functional Area: Catching (FA:C)				
Domain: Manipulative Small Motor (D:MSM)				
Functional Area: Writing (FA:W)				
Functional Area: Copying Designs (FA:CD)				
Functional Area: Dressing (FA:D)				
Domain: Mobile Gross Motor (D:MGM)				
Functional Area: Walking (FA:WL)				
Functional Area: Climbing (FA:CL)				
Functional Area: Jumping (FA:J)				
Total Score[a]				

*Coefficients marked with * have alpha levels greater than .05.
[a]Total score is a summation of all functional area scores.

FA:C	D:MSM	FA:WR	FA:CD	FA:D	D:MGM	FA:WL	FA:CL	FA:J	TOTAL
.42	.07*[1]	.20*	.01*	.28*	.29*	.15*	.20*	.25*	.40
.01*	-.01*	.13*	-.09*	.38	.41	.33*	.10*	.41	.26*
.15*	.01*	.10*	.01*	-.03	.08*	-.03*	.20*	.02*	.25*
.34	.26*	.13*	.25*	.01*	-.11*	.06*	-.21*	-.05*	.30*
	.11*	.21*	.01*	.48	.27*	.06*	.19*	.25*	.27*
		.21*	.98	.09*	.08*	.19*	.14*	.01*	.91
			.06*	.34	.19*	.27*	.05*	.18*	.28*
				-.10	.01*	.13*	.13*	-.08*	.85
					.37	.22*	.05*	.42	.25*
						.64	.50	.93	.42
							.08*	.59	.37
								.18*	.31*
									.33*

Analyses of Variance

To determine if any developmental differences due to age or sex of the children contributed to differences in performance on the motor tasks, 26 regression analyses of variance (ANOVAs) were computed. Analyses of variance tables and regression coefficients for each of the 26 motor tasks which served as dependent variables can be seen in Appendix C.

Effects of age on task performance. Of the 26 ANOVAs, nine produced no significant effects for age at a .05 level. The tasks which served as dependent variables in these analyses were; (1) throwing the large ball, (2) throwing the small ball, (3) throw-catch, (4) ground level kick, (5) drop kick, (6) copying a cross, (7) copying a circle, (8) climbing upstairs, and (9) climbing downstairs. Analyses of the remaining 17 motor tasks indicated a significant main effect for age.

Means for the three age groups in the 17 analyses showing effects for age were further analyzed using Tukey's HSD test for pairwise comparisons (Kirk, 1968). The alpha level for these comparison tests was set at .05. Under the stationary gross motor domain, only three tasks showed overall main effects for age. They were (1) the catch of a bounced ball, (2) the left foot balancing task, and (3) the right foot balancing task. Group means for these three tasks can be seen in Table 22. Significant differences were found between all three group

Table 22. Task score means from the stationary gross motor domain showing effect of age.

	Task		
	Left Foot Balance	Right Foot Balance	Bounce - Catch
5 year olds (n=32)	3.44	3.94	2.58
4 year olds (n=35)	2.91	2.71	2.07
3 year olds (n=16)	1.75	1.69	1.56

means on the right foot balancing task and the bounce-catch task. However, on the left foot balancing task, the mean score for the four year olds was found only to be significantly different from the mean of the three-year-old group.

Under the manipulative small motor domain, age of children produced significant differences in performance on "name" writing; copying a square, triangle, rectangle and rectangular design; and the dressing skills of zipping, snapping, buttoning, lacing, and tying. The group means for these tasks can be seen in Table 23. The four copying tasks, the "name" writing task, and the lacing and tying tasks produced significant differences in mean scores between all three age groups. However, for the zipping, snapping and buttoning tasks, significant differences were found only between the mean scores of the three and four year olds.

The domain of mobile gross motor skills showed overall main effects for age on the board walking task and the three jumping tasks. The mean scores of the three age groups for these tasks can be seen in Table 24. Tukey's HSD test revealed significant differences between mean scores of the three and four year olds for all three tasks. However, only the board walking task showed significant differences between the means of the four and five year olds.

In summary, the majority of the analyses of variance

Table 23. Task score means from the manipulative small motor domain showing effect of age.

	TASK									
	NW	CS	CT	CR	CRD	ZIP	SNAP	BUTTON	LACE	TIE
5 year olds (n=32)	6.16	1.69	1.47	2.56	3.84	1.78	8.00	8.00	7.28	1.97
4 year olds (n=35)	5.20	1.29	0.71	1.89	2.31	1.60	7.89	7.66	4.71	1.14
3 year olds (n=16)	3.69	0.31	0.13	0.37	0.44	1.25	6.75	5.13	1.56	0.19

NW = Name Writing
CS = Copying Square
CT = Copying Triangle
CR = Copying Rectangle
CRD = Copying Rectangular Design

Table 24. Task score means from mobile gross motor domain.

	Task			
	Walking Board	Broad Jump	Right Foot Hop	Left Foot Hop
5 year olds (n=32)	3.81	4.56	3.16	3.06
4 year olds (n=35)	3.50	3.86	2.57	2.57
3 year olds (n=16)	2.09	2.81	1.06	1.00

indicated age differences in performance existed and thus satisfied the skill selection criterion that tasks should represent developmental patterns.

Effects of sex on task performance. The 26 analyses of variance in which task score served as dependent variables also indicated that significant main effects for sex existed. The task scores showing these effects were; (1) climbing upstairs, (2) climbing downstairs, (3) buttoning, (4) tying, (5) throwing a large ball, (6) throwing a small ball.

Group means for these six variables can be seen in Table 25. On all six tasks, male scores were significantly higher than female scores.

Only buttoning showed an overall significant effect on performance for the interaction of age and sex of child. Post hoc analyses revealed significant differences existed only between three year old males and the other five groups. However, caution should be used in interpreting these differences since there were only two male three year olds and both performed minimally. Specifically, these two males only buttoned a total of three buttons between them for a group average of 1.5 buttons. In comparison, three-year-old females buttoned an average of 6.71 buttons. All other group means were higher than the mean of the three year old females.

Table 25. Mean task scores by sex of child.

	Small Ball Throw	Large Ball Throw	Buttoning	Tying	Climbing Upstairs	Climbing Downstairs
Males (n=34)	4.50	3.38	7.50	1.32	1.74	1.56
Females (n=49)	2.98	2.31	7.08	1.24	1.51	1.06

In summary, while analyses indicated sex differences existed, these differences might be attributable to the small number of subjects within groups. Further research is needed before these results are accepted.

CONCLUSIONS

The main thesis of this study was that a new perspective of educational measurement was needed. Specifically, it was argued that NRT and CRT had individually failed to meet the instructional needs identified by Glaser (1963) and others over 15 years ago. Furthermore, it was argued that a review of the educational measurement literature suggested that a measurement approach combining certain NRT and CRT characteristics might provide what was needed. To demonstrate the feasibility of these arguments, a perceptual-motor test was constructed which integrated characteristics of both measurement approaches. In constructing the test, an attempt also was made to address some of the specific motor assessment needs in early childhood education.

The results of this study provided information regarding two essential areas of test construction: (1) the selection and organization of test content, and (2) the design of a reliable and useful scoring system. Under the selection and organization of test content, the results of the study provided information concerning the following questions:

(1) Did the content of the test focus on nontrivial developmental accomplishments?

(2) Were skill domains constructed in such a way that a clear understanding of test content could be

obtained by the test user?

(3) Were the skill domains relatively independent so that instructional programs could focus primarily on the essential characteristics of each domain?

Under the area concerned with the design of a reliable and useful scoring system, information concerning three questions was provided by the results.

(1) Could the scoring system provide information useful to instructional manipulation?

(2) Could the scoring system permit clear communication of results to layment and educators alike?

(3) Did the task scores produce consistent information?

<u>Selection and Organization of Content</u>

Three kinds of information supported the conclusion that the content of the PPMS focused on nontrivial developmental accomplishments. First, the task selection criteria insured that the content represented (1) directly observable behaviors, (2) both fine and gross motor activities, and (3) skills which have been associated with studies of both normal and deviant behavioral development. Second, each of these tasks directly represent or relate to motor competencies which will be required over the life span.

The last group of information supported the conclusion that the content of the PPMS was generally developmental in nature. On 17 of the 26 analyses of variance where task scores were the dependent variable, significant effects for age were found. While analyses of nine task scores produced no such results, this did not necessarily mean that development of these skills occur earlier or later than the age span examined in this test. Since information in the literature concerning the development of many of these areas was sparce, scoring may not have reflected the essential developmental aspects of performance. Future research into the qualitative and quantitative aspects of skill development in these areas might help focus future assessment practices. In the meantime, adjustments in scoring and administrative procedures might be tried to determine how they may have influenced the scoring of these tasks.

The results of the analyses of variance indicated that the content of the test did reflect some sex differences. In particular, males scored higher on climbing up and down stairs, buttoning, tying, and throwing a large and small ball. While previous sex differences have been reported for throwing showing boys better in throwing accuracy, distance and development of efficient overarm throwing patterns, these differences don't usually appear until five years of age (e.g., Seils, 1951). No data could be found regarding sex differences on the buttoning and tying tasks. Both Cratty (1970) and Espenschade and Eckert (1967) did not report any sex difference during preschool years in stair climbing ability. In fact, few sex differences before age five in motor skills have been reported. Further research is needed to verify the present findings before content or scoring adjustments are made.

Organization of PPMS content. The motor tasks of the PPMS were organized so that the test user could clearly understand what behaviors were being measured. Tasks with common performance characteristics were organized under a functional area, which in turn were placed under an appropriate motor domain. A comprehensive written description of each domain was provided which carefully stipulated domain parameters and limitations. Concrete examples of representative behaviors were included in the description to emphasize generalizability of domain content, and to remind the test user that the behaviors tested in the PPMS were only samples of a larger domain.

The organization of motor skills and the written domain descriptions should assist the classroom teacher in selecting similar behaviors for daily practice and testing. Without these characteristics the teacher would have only a vague understanding of the behaviors being assessed. Therefore, when the test indicated that a student performed poorly, the teacher would be left to either guessing which other behaviors could be used for remedial work or teaching to the specific items which indicated poor performance. In either case, the probability of nontransferability of training would be greater than if clear content descriptions had been provided.

The test administrator, as well as the educator, should benefit from the organization of the PPMS and the written descriptions it includes. The grouping of behaviors with common performance characteristics should focus the attention of the tester and assist in the observing and scoring of behavior. These organizational characteristics should be particularly helpful since qualitative aspects of motor performance were emphasized on several tasks.

Independence of motor domains. The results of the factor analysis and the correlation of functional area and domain scores indicated that the set of motor skills included in the PPMS can not be grouped to provide relatively distinct indexes of motor functioning. Specifically, correlation coefficients among functional areas within domains often were below .20 and even at times were negative. Furthermore, the correlation coefficients among functional areas across domains were not lower than those within domains.

The results of the factor analysis did not provide empirical support for the organization of content into separate components. One factor resulted from this analysis, and it sufficiently accounted for the variation in performance on most items. Obviously, the content of the PPMS was not heterogeneous enough to warrant separate performance indexes. Rather, performance across tasks was highly related and reflected more a single general developmental motor ability, rather than several separate skill areas.

Further refinement of the test's content is needed so that empirically distinct domains can be organized. More information concerning all aspects of motor development and movement is needed to assist these efforts. Unfortunately, as Teeple (1978) and Roberton (1978) have stated, information concerning motor performance of preschool-age children is often scattered or unfocused. Until that information is available, educators and measurement specialists will need to work together to organize test content as logically as possible.

The Scoring System

Reliability of scoring. Three aspects of reliability were considered in the present study: (1) inter-rater reliabilities of scoring procedures for the copying of geometric designs, (2) internal consistency estimates, and (3) stability estimates. The coefficients produced by the analysis of ratings of two independent judgements while scoring the copying of designs indicated scoring criteria could be used with a fairly high degree of consistency. However, since observational procedures and not direct measurements of behavior were often used when collecting qualitative data, further studies will be needed to determine whether adequate consistent agreement exists for all observational procedures.

Internal consistency coefficients indicated a fairly high degree of consistency across all items for each age group. These data combined with the results of the factor analysis supported the conclusion that the PPMS is fairly homogeneous in content.

Test-retest reliability data indicated that some task scores were very reliable over time and others were not. For example, "name" writing showed a very high consistency for all three age groups, however, the throw-catch task scores were shown to be very unreliable for four year olds. While domain scores were generally more reliable than task scores, improvement is still needed before the PPMS can be used for individual assessment. Studies, exploring possible ways to improve the stability of the scoring over time, should investigate how much of this variability in performance across time is due to administrative or scoring procedures; and how much is attributable to developmental variability within the subject.

A clear examination of process scores between testing occasions indicated that in some cases scoring reflected subtle variance in developmental growth patterns. In particular, when more than two developmental levels of performance were measured, the children moved to an adjacent level rather than skipping levels. For example, when the process score on the throw-catch task changed on the second testing occasion, every time the change reflected a movement from Wellman's (1937) developmental level one to level two or vice versa; never from level one to three or vice versa. Similar changes in scores were found on the small-ball throw, particularly in the three-year-old age group.

Instructional utility. Since qualitative and quantitative aspects of movement were integrated into the scoring system, the educator would have information which could be used to plan instructional programs. For example, a teacher using the scoring system of the PPMS would know more than whether or not a ball was caught, or how many catches were made on successive throws. The teacher also would know what particular movements were used in performing the task.

While number of successful catches may allow normative comparisons and provide information regarding whether or not a child is developmentally delayed, it provides little instructionally useful information. Why did the child miss the ball? Was s/he at a particular level of development where little effort is made to move the body to adjust to the flight of the ball? Also what arm movements did the child make in trying to catch the ball? Since this kind of information was included in the PPMS scoring system, the educator familiar with the development of throwing would know where to begin in planning remedial efforts. Also the educator would have information for evaluating the remedial efforts. Specifically, since s/he knew where the child was developmentally, s/he would also know what particular changes in behavior must occur if the program was to be judged successful.

The scoring system of the PPMS also should offer a logical alternative to the instructional programs which were previously based solely on quantitative scores. For example, a remedial program for a child who was found to be significantly delayed in catching a tossed ball might specify an objective of catching a ball on eight of ten tosses. Since this instructional program does not include qualitative information, the instruction may be prematurely discontinued when the limited objective has been achieved. Halverson (1971) demonstrated that we can not assume that a level of development has been reached because certain goals have been achieved. Unless we are aware of the developmental sequences, or the process aspects of movement, making such assumptions at one point in a child's life may lead to frustration and failure at another point when more complex movements are required. Halverson further stated:

> Can we assess progress of a child in movement if we can not learn to see what he is doing in movement? Can we learn to provide a rich and varied movement experience for a child if we never learn to see if, in fact, he is everyvarying

> his movement? Can we afford to ignore whether he is learning to move with increasing competence within the environment in which we put him? I think not (p. 29).

Communication of results. The scoring system of the PPMS was designed to provide normative comparisons. When the final version of the PPMS is completed, it will be administered to a large number of children who will serve as the reference population. This sample will be defined in terms of age, sex, socioeconomic status and geographic location. As in this experimental version, individual task scores will be summed to provide functional area and domain scores. The domain score will serve as a general developmental index, while the functional area score will serve as a more specific motor skill index. If a child is found to be performing significantly below age level, the educator can examine the product and process scores of each functional area to determine why the particular deficit exists. The domain score will provide the more familiar normative reference for communicating results to the lay person that Hively (1974) and Popham (1976) recommended.

Final remarks. In summary, the results of this study have generally indicated that combining CRT and NRT characteristics is a logical measurement alternative to meet current instructional needs. While further research is needed to improve the organization of content into domains and refine the scoring system, the general feasibility of such an approach has been demonstrated. Specifically, the content of the test focused on nontrivial competencies and reflected a comprehensive set of needs within preschool motor assessment. Furthermore, the test included a clear description of its content which would facilitate instructional practices.

The scoring system of the PPMS combined both normative and criterion references which provided information usable in instructional situations. This combination of scoring characteristics also should improve the interpretation and communication of results to the lay public and the educator.

Very recently, other individuals within educational measurement are independently recognizing the feasibility of the measurement approach presented in this paper. Specifically, Bank and Burry (1978) have described future objectives of the Center of the Study of Evaluation (CSE) at the University of California at Los Angeles. They outlined a need within educational measurement similar to the thesis of this paper.

> A theory of test content with corresponding qualitative and quantitative modes needs to be explored. Without such a theory, systematic improvement of the quality of achievement tests is unlikely. Researchers in measurement theory seem to limit their study to decision-theoretic approaches and questions of test reliability using alternative statistical models. This kind of inquiry is extremely important and will be continued in the program.

> However, relatively little work has been conducted in other aspects of theory development, such as design characteristics of tests (p. 5).

With such attempts underway and with continued efforts which reflect a willingness to modify and combine various measurement approaches, educational testing theory can progress to meet instructional challenges by providing the needed assessment devices.

REFERENCE NOTES

1. Davis, F. B. Criterion-referenced tests: A critique. Paper presented at the meeting of the American Educational Research Association, New York, February, 1971 (ERIC Document Reproduction Service No. ED 050 154, 11 pp.)

2. Fuqua, R. W. & Phye, G. D. The use of a behavioral developmental profile in the evaluation of a rural home intervention program for handicapped preschoolers. A paper presented at the annual meeting of the American Educational Research Association, New York, April, 1978.

3. Huynh, H. Statistical problems in binary classification. Paper presented at the annual meeting of the American Statistical Association, Atlanta, August, 1975.

4. Block, J. H. Student evaluation: Toward the setting of mastery performance standards. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, April, 1972.

5. Kriewall, T. E. Applications of information theory and acceptance sampling principles to the management of mathematics instruction. Unpublished doctoral dissertation, University of Wisconsin, 1969.

6. Ivens, S. H. An investigation of item analysis, reliability and validity in relation to criterion-referenced tests. Unpublished doctoral dissertation, Florida State University, 1970.

7. Johnson, T. J. Evaluation in the context of product and program development in laboratory and research and development centers. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April, 1972.

8. Woodson, M.I.C.E. Classical test theory and criterion-referenced scales. (ERIC Document Reproduction Service No. ED 083 298, 13 pp.), 1975.

9. Loovis, E. M. Model for individualizing physical education experiences for the preschool moderately retarded child. (Doctoral dissertation, the Ohio State University, 1975).

Dissertation Abstracts International, 1976, _36_:5126A (University Microfilms No. 76-3485).

10. Nelson, D. C. Motor ability and motor activity in young children. Unpublished doctoral dissertation, Iowa State University, 1971.

REFERENCES

Anderson, R., Miles, M., & Matheny, P. Communicative evaluation chart. Cambridge, Mass.: Educators Publishing Service, Inc., 1963.

Andrews, B. J., & Hecht, J. T. A preliminary investigation of two procedures for setting examination standards. Educational and Psychological Measurement, 1976, 36, 45-50.

Baker, E. L. Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Bank, A. & Burry, J. Directions and testing evaluation: CSE, 1978. Evaluation Comment, 1978, 5, 1-9.

Beatty, L. S., Madden, R., & Gardner, E. F. Stanford diagnostic mathematics test. New York: Harcourt Brace Jovanovich, in press.

Bender, L. Bender motor gestalt test. Albany, N. Y.: American Orthopsychiatric Association, 1938.

Block, J. H. Criterion-referenced measurements: Potential. School Review, 1971, 69, 289-298.

Bloom, B. S., Hastings, S. T., & Madaus, G. F. Handbook of formative and summative evaluation of student learning. New York: McGraw-Hill, 1972.

Brigance, A. H. Brigance diagnostic inventory of early development. Woburn, Mass.: Curriculum Associates, Inc., 1978.

Bronfenbrenner, U. Is early intervention effective. In C. L. Streuning and M. Guttentag (Eds.) Handbook of evaluation research. Beverly Hills: Sage Publications, 1975.

Carroll, J. B. Problems of measurement related to the concept of learning for mastery. Educational Horizons, 1970, 48, 71-80.

Carver, R. P. Special problems in measuring change with psychometric devices. In _Evaluation research: Strategies and methods_. Washington: American Institutes of Research, 1970.

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. _American Psychologist_, 1974, _29_, 512-518.

Charles, D. C. Historical antecedents of life-span developmental psychology. In L. R. Goulet and P. B. Baltes (Eds.), _Life-span developmental psychology: Research and theory_. New York: Academic Press, 1970.

Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement of partial credit. _Psychological Bulletin_, 1968, _70_, 213-220.

Cratty, B. J. _Perceptual and motor development in infants and children_. New York: The Macmillan Co., 1970.

Cratty, B. J. & Martin, M. M. Perceptual motor efficiency in children. _Journal of Applied Psychology_ _40_, 1956, 2.

Cronbach, L., Rajaratnam, N., Gleser, G. Theory of generalizability: A liberalization of reliability theory. _British Journal of Statistical Psychology_, 1963, _16_, 137-163.

Cross, L. & Goins, K. (Eds.) _Identifying handicapped children: A guide to casefinding, screening, diagnosis, assessment and evaluation_. New York: Walker and Company, 1977.

Denham, C. H. Criterion-referenced, domain-referenced and norm-referenced measurement: A parallax view. _Educational Technology_, 1975, _25_, 9-13.

DeOreo, K. L. _DeOreo fundamental motor skills inventory_. Kent, Ohio: Kent State University, 1976.

Dickerson, J., Evanson, M. & Spurlock, L. _Pupil progress evaluation plan_. Coeur d'Alene, ID.: Panhandle Child Development Association, Inc., 1975.

Donahue, M., Montgomery, J. D., Keiser, A. F., Roecher, V. L., & Smith, L. I., _Marshalltown behavioral developmental profile_. Marshalltown, Ia.: Area Education Agency 6, 1975.

Duncan, A. D. Tracking behavioral growth: Day to day measures of frequency over domains of performance. In W. Hively (Ed.) _Domain-referenced testing_. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Ebel, R. L. Criterion-referenced measurement: Limitations. _School Review_, 1971, _79_, 282-288.

Ebel, R. L. Evaluation and educational objectives. _Journal of Educational Measurement_, 1973, _10_, 273-279.

Espenschade, A.S. & Eckert, H. M. _Motor development_. Columbus, Ohio: Charles E. Merrill Books, Inc., 1967.

Frostig, M. _Frostig developmental test of visual perception_. Palo Alto, Calif.: Consulting Psychologists Press, 1961.

Gesell, A., & Thompson, H. _The psychology of early growth._ New York: Macmillan, 1938.

Getman, G. N. _How to develop your child's intelligence, a research publication_. Luverne, Minn.: G. N. Getman, 1952.

Glaser, G. R. Instructional technology and the measurement of learning outcomes. _American Psychologist_, 1963, _18_, 519-521.

Glaser, G. R. Components of a psychology of instruction: Toward a science of design. _Review of Educational Research_, 1976, _46_, 1-24.

Glaser, G. R., & Nitko, A. J. Measurement in learning and instruction. In R. L. Thorndike (Ed.), _Educational Measurement_ (2nd ed.). Washington, D. C.: American Council on Education, 1971, 625-670.

Glass, G. V., & Smith, M. L. The technology and politics of standards. _Educational Technology_, 1978, _28_, 12-18.

Gray, W. M. A comparison of Piagetian theory and criterion-referenced measurement. _Review of Educational Research_, 1978, _48_, 223-249.

Guttridge, M. V. A study of motor achievements of young children. _Archives for Psychologie_, 1939, 244, 1-178.

Halverson, L. E. The significance of motor development. In G. Engstrom (Ed.) _The significance of the young child's motor development_. Washington, D. C.: National Association for the Education of Young Children, 1971.

Halverson, H. M. An experimental study of prehension in infants by means of systematic cinema records. _Genetic Psychology Monograph_, 1931, _10_, 107-286.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. _Journal of Educational Measurement_, 1973, _10_, 159-170.

Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. B. Criterion-referenced testing and measurement: A review of technical issues and developments. _Review of Educational Research_, 1978, _48_, 1-47.

Hellebrandt, F. A., Rarich, G. L., Glassow, R., & Carns, M. L. Physiological analysis of basic motor skills. 1. Growth and development of jumping. _American Journal of Physical Medicine_, 1961, _40_, 14-25.

Hentschke, G. C., & Levine, D. M. Planning for evaluation in performance contracting experiments: The connection to domain-referenced testing theory. In W. Hively (Ed.) _Domain-referenced testing_. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Herkowitz, J. Assessing the motor development of children: Presentation and critique of tests. In M. V. Ridenour (Ed.) _Motor development: Issues and applications_. Princeton, N. J.: Princeton Book Co., Publishers, 1978.

Hildreth, G. Developmental sequences in name writing. _Child Development_, 1936, _7_, 291-302.

Hively, W. Introduction to domain-referenced testing. In W. Hively (Ed.) _Domain-referenced testing_. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Hively, W., Maxwell, M. G., Rabehl, G. J., Sension, D. B., & Lundin, S. _Domain-referenced curriculum evaluation_. Los Angeles, Calif.: Center for the Study of Evaluation, University of California, 1973.

Hively, W., Patterson, H. L., & Page, S. A. A "universe-defined" system of arithmetic achievement tests. Journal of Educational Measurement, 1968, 5, 275-290.

Hoffman, B. The tyranny of testing. London: Collier-Macmillan, 1964.

Huynh, H. On the reliability of decisions in domain-referenced testing. Journal of Educational Measurement, 1976, 13, 253-264.

Johnson, O. G. Tests and measurements in child development: Handbook II. Washington: Jossey-Bass Publishers, 1976.

Johnson, O. G., & Bommarito, J. W. Tests and measurements in child development: Handbook I. Washington, D. C.: Jossey-Bass Publishers, 1974.

Johnson, T. J. Program and product evaluation from a domain-referenced viewpoint. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Karlsen, B., Madden, R., & Gardner, E. F. Stanford diagnostic reading test. New York: Harcourt Brace Jovanovich, 1977.

Kephart, N. C. Perceptual-motor aspects of learning disabilities. Exceptional Children, 1964, 31, 201-206.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks/Cole Publishing Co., 1968.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Madden, R., Gardner, E. R., Rudman, H. C., Karlsen, B., & Mervin, J. C. Stanford achievement test. New York: Harcourt Brace Jovanovich, 1973.

Mager, R. F. Preparing instructional objectives. Palo Alto: Fearon Publishers, 1962.

McGraw, M. B. Growth: A study of Johnny and Jimmy. New York: Appleton-Century, 1935.

Meisels, S. J., Wiske, M. S. Eliot-Pearson screening profile. Medford, Mass.: Tufts University, 1976.

Meskauskas, J. A. Evaluation models for criterion-referenced testing: Views regarding mastery and standard-setting. Review of Educational Research, 1976, 46, 133-158.

Messick, S. A. The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 1975, 30, 955-966.

Miller, D. M. Content, items, decisions: Orienting curriculum assessment surveys to curriculum management and modification. In W. Hively (Ed.) Domain-referenced testing. Englewood, Cliffs, N. J.: Educational Technology Publications, 1974.

Millman, J. Passing scores and test length for domain-referenced measures. Review of Educational Research, 1973, 43, 205-216.

Millman, J. Program assessment, criterion-referenced tests, and things like that. Educational Horizons, 1974, 52, 188-192. (a)

Millman, J. Sampling domains for domain-referenced tests. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliff, N. J.: Educational Technology Publication, 1974. (b)

Millman, J., & Popham, W. J. The issue of item and test variance for criterion-referenced tests: A clarification. Journal of Educational Measurement, 1974, 11, 137-138.

Nitko, A. J., & Hsu, T. Using domain-referenced tests for student placement, diagnosis and attainment in a system of adaptive, individualized instruction. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliff, N. J.: Educational Technology Publications, 1974.

Nunnally, J. C. Introduction to psychological measurement. New York: McGraw-Hill, 1967.

Popham, W. J. Selecting objectives and generating test items for objectives-based tests. Los Angeles: Instructional Objectives Exchange, 1972.

Popham, W. J. Establishing performance standards. Englewood Cliffs, N. J.: Prentice-Hall, 1973.

Popham, W. J. Teacher evaluation and domain-referenced measurement. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Popham, W. J. Educational Evaluation. Englewood Cliffs, N. J.: Prentice-Hall, 1975.

Popham, W. J. Normative data for criterion-referenced tests? Phi Delta Kappan, 1976, 58, 593-594.

Popham, W. J., & Husek, T. R. Implications for criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Roberton, M. A. Stages in motor development. In M. V. Ridenous (Ed.) Motor development: Issues and applications. Princeton: Princeton Book Co., 1978.

Ryan, S. (Ed.) A report on longitudinal evaluations of preschool programs. Washington: Office of Child Development, 1972.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education. Boston: Houghton Mifflin Co., 1978.

Sanford, H. Learning accomplishment profile. Winston-Salem, N. C.: Kaplan School Supply, 1975.

Seils, L. The relationship between measure of physical growth and gross motor performance of primary-grade children. Research Quarterly, 1951, 22, 244-260.

Sension, D. B., & Rabehl, G. J. Test-item domains and instructional accountability. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliffs, N. J.: Educational Technology Publications, 1974.

Sharp, E. Assessment by behavior rating. Tuscon: University of Arizona, 1975.

Shirley, M. M. The first two years: A study of twenty-five babies. Volume I. Postural and locomotor development. Minneapolis: University of Minnesota Press, 1931.

Subkoviak, M. J., & Baker, F. B. Test theory. In L. S. Shulman (Ed.) Review of Research in Education. Itasca, Ill.: F. E. Peacock Publishers, Inc., 1977.

Swaminathan, H., Hambleton, R. K., & Algina, J. A decision theoretic approach to issues in criterion-referenced assessment. Laboratory of Psychometric and Evaluative Research Report No. 1. Amherst, Mass.: School of Education, University of Massachusetts, 1973.

Swaminathan, H., Hambleton, R. K., & Algina, J. Reliability of criterion-referenced tests: A decision-theoretic formulation. Journal of Educational Measurement, 1974, 11, 263-268.

Tallmadge, G. K., & Horst, D. P. A procedural guide for validating achievement gains in educational projects. Washington, D. C.: Government Printing Office, 1976.

Teeple, J. Physical growth and maturation. In M. V. Ridenour (Ed.) Motor development: Issues and applications, Princeton: Princeton Book Co., 1978.

Umansky, W. Assessment-Programming guide for infants and preschoolers. Columbus, Ind.: Developmental Services, Inc., 1974.

Wellman, B. L. Motor achievements of preschool children. Childhood Education, 1937, 13, 311-316.

Wendt, R., Schramm, R., & Schmaltz, D. Early childhood assessment: A criterion-referenced screening device. Waupon, Wisc.: Cooperative Educational Service Agency 13, Ca., 1975.

Wild, M. R. The behavior pattern of throwing and some observations concerning its course of development in children. Research Quarterly, 1938, 9, 20-24.

Woodson, M.I.C.E. The issue of item and test variance for criterion-referenced tests. Journal of Educational Measurement, 1974, 11, 63-64. (a)

Woodson, M.I.C.E. The issue of item and test variance for criterion-referenced tests: A reply. Journal of Educational Measurement, 1974, 11, 139-140. (b)

Ysseldyke, J. E., & Salvia, J. Diagnostic-prescriptive teaching: Two models. Exceptional Children, 1974, 40, 181-185.

APPENDIX A:

PRESCHOOL MOTOR ASSESSMENT BIBLIOGRAPHY

Preschool Motor Assessment Bibliography

I. Standardized instruments

A. Norm-referenced

1. Bender Motor Gestalt Test (L. Bender) Albany: American Orthopsychiatric Association, 1938.

2. Comprehensive Identification Process (R. R. Zehrbach). Bensenville: Scholastic Testing Service, 1975.

3. Denver Developmental Screening Test. (W. K. Frankenburg, J. B. Dobbs, & A. W. Fandal) Denver: Ladoca Project and Publishing Foundation, 1970.

4. Developmental Indicators for the Assessment of Learning (C. Mardell, D. Goldenberg). Highland Park, Ill.: DIAL Inc., 1975.

5. Gesell Developmental Schedules (A. Gesell and C. S. Amatruda) New York: Psychological Company, 1949.

6. McCarthy Scales of Children's Abilities (D. McCarthy). Yew York: The Psychological Corporation, 1972.

7. Modified Lincoln-Oseretsky Motor Development Scale (I. Bailer, L. Doll, and B. G. Winsberg) New York: New York State Department of Mental Hygiene, 1973.

8. Preschool Screening System: Start of a Longitudinal Preventive Approach. (P. Hainsworth and M. Hainsworth). Pawtucket: First Step Publications, 1974.

B. Criterion-referenced

1. Callier-Azusa Scales (R. Stillman, ed.) Dallas: Callier Center for Communication

2. The Lexington Developmental Scales (J. Irwin, C.A. Coleman). Lexington: Child Development Centers, Inc., 1974.

II. Nonstandardized instruments

A. Norm-referenced

1. ABC Inventory (N. Adair and G. Blesch). Muskegan, MI: Research Concepts, 1965.

2. Berry-Buktenica Developmental Test of Visual-motor Integration(K. Berry and N. Buktenica). Chicago: Follett Educational Corporation, 1967.

3. Communicative Evaluation Chart (R. Anderson, M. Miles, and P. Matheny). Cambridge: Educators Publishing Service, Inc.,1963.

4. Comprehensive Early Education Profile (K. F. King, L. Joyner, M. Smith, A. Thompson, M. Meredith, L. Bishop, J. Newquist, L. Brewer, Q. I. Davis). Birmingham: Comprehensive Early Education Program, 1975.

5. Developmental Profile (G. Alpern and T. Boll). Indianapolis: Psychological Development Publications, 1972.

6. Eliot-Pearson Screening Profile (S. J. Meisels, M. S. Wiske). Medford: Tufts University, 1976.

7. The Functional Profile. Peoria: Peoria Association for Retarded Citizens and United Cerebral Palsy of Peoria, 1974.

8. Oseretsky Test of Motor Proficiency (E. A. Doll, ed.). Circle Pines, MN: American Guidance Service, Inc., 1946.

9. Preschool Attainment Record (E. A. Doll). Circle Pines, MN: American Guidance

Service, Inc. 1966.

10. SEED Reflex and Therapeutic Evaluation (G. Jorgenson, S. Wolfe, and P. Pollan). Denver: Sewall Rehabilitation Center, 1975.

11. Vallett Developmental Survey (R.E. Vallett). Palo Alto: Consulting Psychologists Press, 1966.

B. Criterion-referenced

1. Assessment by Behavior Rating (E. Sharp). Tucson: University of Arizona, 1975.

2. Assessment-Programming Guide for Infants and Preschoolers (W. Umansky). Columbus: Developmental Services, Inc. 1974.

3. Brigance Diagnostic Inventory of Early Development (A.H. Brigance). Woburn, Mass.: Curriculum Associates, Inc., 1978.

4. Carolina Developmental Profile (D. L. Lillie). Winston-Salem, N.C.: Kaplan School Supply, 1976.

5. Classroom Screening (N. Giessman, S. Hering, S. Issacson, A. Fazie, and C. Tarchin). Piedmont: Circle Preschool, 1975.

6. Criterion-Referenced Placement Tests. Logan, UT: Mapps Project, 1975.

7. Early Childhood Assessment: A Criterion Referenced Screening Device (R. Wendt, R. Schramm, D. Schmaltz). Waupon, Wisc.: Cooperative Educational Service Agency 13, Ca., 1975.

8. Individual Child Assessment (S. Hering, A. Fazio, and J. Hailey). Piedmont: Circle Preschool, 1975.

9. Informal Teacher Assessment Instrument (W. Drezek). Austin: Infant-Parent Training Program, 1975.

10. Koontz Child Developmental Program: Training Activities for the First 48 Months (C. Koontz). Los Angeles: Western Psychological Services, 1974.

11. Learning Accomplishment Profile (A. Sanford). Winston-Salem, N.C.: Kaplan School Supply, 1975.

12. Learning Accomplishment Profile-Diagnostic Edition (B. M. Griffen, A. R. Sanford, and D. C. Wilson). Winston Salem, N.C.: Kaplan School Supply, 1975.

13. Marshalltown Behavioral Development Profile (M. Donahue, J. Montgomery, and others). Marshalltown, Ia: AEA #6 Preschool Division, 1975.

14. Memphis Comprehensive Development Scales (A. P. Quick, T. L. Little, A. A. Campbell). Belmont: Fearon Publishers, 1974.

15. Move-Grow-Learn Survey (R. E. Orpet and L. L. Huestis). Chicago: Follett Publishing Company, 1967.

16. Portage Guide to Early Education, revised edition (S. Bluma, M. Shearer, A. Frohman, and J. Hilliard). Portage: Portage Project, 1976.

17. Pupil Progress Evaluation Plan (J. Dickerson, M. Evanson, and L. Spurlock). Coeur d'Alene, ID: Panhandle Child Development Association Inc., 1975.

18. SEEC Developmental Wheel (J. E. Swanson and Staff). Schaumburg: Early Childhood Education, 1976.

19. The Teaching Research Placement Test, in the Teaching Research Curriculum for Moderately and Severely Handicapped (H. D. Fredricks et al.). Springfield: Charles C. Thomas, 1976.

III. Experimental instruments

1. DeOreo Fundamental Motor Skills Inventory (K. L. DeOreo). Kent, Ohio: Kent State University, 1976.

2. Ohio State University Scale of Intra-Gross Motor Assessment (E.M. Loovis) Columbus, Ohio: Ohio State University, 1975.

APPENDIX B:

PRESCHOOL PERCEPTUAL MOTOR SCALE

Preschool Perceptual-Motor Scale

I. Stationary Gross Motor Domain

Domain description and limitations. The content of this domain should include the performance of gross motor behaviors performed in a relatively stationary position. Movements primarily involving one or both arms, one or both legs, or coordination of arm(s) and leg(s), or movements generally involving large muscle activity. However, these movements should be performed in a restricted area of space. That is, they should not require the body to be moved more than two steps in any direction. At least one behavior must include the use of some equipment. However, at least one behavior also should not involve equipment. The equipment can be thrown, or used to strike other equipment, or can be caught by the individual. Behaviors which meet domain specifications and could be included in a domain sample, but not to which a domain sample is restricted are:

1. Variations of stationary balancing, e.g.,
   - (a) balancing on single foot; eyes closed.
   - (b) balancing on tip toe; eyes opened.
   - (c) balancing on single foot; arms

crossed on chest.

(d) balancing on tiptoe; arms extended and eyes closed.

2. Rising from seated position on floor; legs crossed; arms extended.
3. Cross-legged sit to floor from standing position, arms crossed on chest.
4. Throwing of objects.
5. Striking an object with another, e.g., striking thrown ball with bat; striking stationary ball with bat.
6. Kicking an object, e.g., drop kick ball; stationary kick of ball.
7. Catching object, e.g., catching thrown ball; catching rolling ball; catching bounced ball.

A. Functional area: Balancing

1. Left foot lifted (stork position).

<u>Equipment</u>: None

<u>Directions</u>: Child may hold arms in any position. Duration of the demonstration should be timed using a stop watch. Begin timing when foot clears the floor and stop timing as soon as any part of the foot touches the floor.

Record duration on record form to nearest tenth of a second. The child may hold onto an object such as the wall or a chair if s/he is 30 months old or younger. However, the examiner should <u>not</u> physically assist the child, i.e., the child is not allowed to hold on to the examiner. The child should stand flat footed with eyes open. Examiner says: "Please lift this foot off the ground." (Touches child's left foot) "Good. Now I want to see how long you can stand on one foot. When I say 'Go' lift this foot off the ground." (Touch child's left foot) "Go." Begin timing. Record time and whether assistance was used.

2. Right foot lifted (stork position).

<u>Equipment</u>: None

<u>Directions</u>: Administer as in Item 1 except touch child's right foot before

saying "Go".

B. Functional Area: Throwing

1. Throwing small ball.

<u>Equipment</u>: Rubber ball: eight inches in circumference.

<u>Directions:</u> Give the ball to the child and encourage her/him to throw it as far as s/he can. The examiner should observe and record the manner in which the child throws the ball, noticing and recording:

(a) whether overhand or an underhand position was used;

(b) whether or not the child steps with a foot while throwing;

(c) whether the foot used is on the same or opposite side from which the ball is thrown.

Finally, the distance of the throw should be measured and recorded for three trials.

2. Throwing a large ball.

<u>Equipment</u>: Rubber ball: twenty inches in circumference.

Directions: Administer as in item 1. Observe and record the same data on three trials.

C. Functional Area: Kicking

1. Ground level kick.

Equipment: Rubber ball: 20 inches in circumference.

Directions: Demonstrate kicking the ball at ground level, stationary position. Place the ball in front of the child and encourage her/him to kick it. Record whether or not the child kicks the ball in a forward direction on each of three trials.

2. Drop kick.

Equipment: Rubber ball: 20 inches in circumference.

Directions: Demonstrate standing stationary position, holding the large ball between your hands, dropping the ball and kicking it forward. Encourage the child to do the same. Administer 3 trials. Record whether or not the ball was kicked on each trial.

D. Functional Area: Catching

1. Catching a large ball thrown.

Equipment: Rubber ball: 20 inches in circumference.

Directions: Stand approximately 3 to 4 feet in front of the child and encourage the child to catch the ball. Throw ball at child's sternum. Besides recording whether or not the ball is caught, observe the arm position used by the child. Record if:

(a) a arm-cradle position,

(b) a stiff arm position or,

(c) a relaxed arm positions is used.

Administer 3 trials.

2. Catching a large ball bounced.

Equipment: Rubber ball: 20 inches in circumference.

Directions: Stand approximately 10 to 12 feet from the child and bounce the large ball between the examiner and the child so that it arrives chest high to the child. Administer 3 trials. Record number of catches.

II. Manipulative Small Motor Domain

Domain description and limitations. This domain includes the performance of manipulative small motor movements. Behavioral content includes manipulation of eyes, fingers, and hands, or the coordination of a combination of these features. This does not exclude behaviors in which some arm movement is secondarily necessary to perform a specific small motor skill. For example, the coordination of eyes, hands and fingers in the threading of beads involves the movement of the arm picking up and threading of the beads. However, primary interest should focus on the use of the eyes, hands and fingers, and their manipulative movements. A skill involving the use of a writing instrument in a coordinated eye-hand movement must be included. For example, writing, prewriting, scribbling, drawing, or copying would meet this specification. Other manipulative equipment such as beads, small blocks, clothing items, paper and scissors may be used, but are not mandatory. Examples of possible domain content are:

(1) bead stringing,

(2) use of scissors in cutting of paper,

(3) copying designs or geometric figures,

(4) dressing skills, e.g.,

(a) snapping,

(b) buttoning,

(5) stacking small blocks or cups,

(6) writing/scribbling,

(7) drawing a figure, e.g., drawing a man.

A. Functional Area: Writing

1. "Name" writing.

<u>Equipment</u>: Pencil or pen and a 8½" X 11" blank sheet of paper.

<u>Directions</u>: Examiner says: "I want you to write your first name for me. Write it here as best you can (point to paper while handing the child a pencil)." If the child indicates that he can't write say, "Let's pretend you can write it. Go ahead. Write it here." Observe how the child holds the pencil.

B. Functional Area: Copying

1. Copying a cross.

<u>Equipment</u>: A pencil and a 8½" X 11" paper divided into 4 quadrants, 3 blank

and 1 containing 2 lines 5 inches long intersecting so one is horizontal and the other perpendicular.

Directions: Show the child the drawing of the cross. Say, "This is a cross. I want you to draw one just like it right here." Administer 3 trials.

2. Copying a circle.

Equipment: Same as item 1 except use a drawing of a circle 4½" in diameter.

Directions: Administer as in item 1 except use circle figure. Administer 3 trials.

3. Copying a square.

Equipment: Same as item 1 except use a drawing of a square with 4" sides.

Directions: Administer as in item 1 except use square figure. Administer 3 trials.

4. Copying a triangle

Equipment: Same as item 1 except using the drawing of a triangle with equal sides measuring 5 inches.

5. Copying a rectangle.

<u>Equipment</u>: Same as item 1 except use a drawing of a rectangle 3" X 6".

<u>Directions</u>: Administer as in item 1 except use the rectangle figure. Administer 3 trials.

6. Copying a rectangle with diagonal line.

<u>Equipment</u>: Same as item 1 except use a drawing of a 3" X 6" rectangle with a dissecting line from lower left corner to upper right corner.

<u>Directions</u>: Administer as in item 1 except use the rectangle with diagonal line figure. Administer 3 trials.

C. Functional Area: Dressing

1. Zipping.

a. Zipping without latching.

<u>Equipment</u>: Zipping dressing frame

<u>Directions</u>: The examiner should first demonstrate zipping up and down the dressing frame. With the zipper latched

at the bottom, encourage the child to do the same. Record whether the child completely zips up and completely zips down. The child need not unlatch the zipper.

b. Latching and zipping.

Directions: The examiner should demonstrate latching the zipper and zip it completely up and down, and then unlatch the zipper. Encourage the child to do the same. The examiner may verbally cue the child if s/he forgets any part of the task. Administer 1 trial.

2. Buttoning.

a. Unbuttoning.

Equipment: Dressing frame with 4 buttons.

Directions: The examiner should demonstrate unbuttoning the dressing frame and then rebutton it. Encourage the child to unbutton the frame. Record the number of buttons successfully unbuttoned. Discontinue after 1½ minutes.

b. Buttoning.

<u>Directions</u>: The examiner should demonstrate buttoning the dressing frame and then encourage the child to do the same. Record the number of buttons buttoned and the time required to complete the task. Discontinue after 2 minutes.

3. Snapping.

a. Unsnapping.

<u>Equipment</u>: Snapping frame with four snaps.

<u>Directions</u>: Administer as item 2a except using snapping frame.

b. Snapping.

<u>Directions</u>: Administer as item 2b except use snapping frame.

4. Lacing.

<u>Equipment</u>: Lacing dressing frame with 5 pairs of eyelets.

<u>Directions</u>: The examiner should demonstrate lacing the frame and then encourage the child to do the same. Record the number of correct placements. Discontinue after 1½ minutes.

5. Tying.

<u>Equipment</u>: Lacing dressing frame with 5 pairs of eyelets.

<u>Directions</u>: The examiner should demonstrate tying a complete shoe tie with 2 bows using the lacing frame. Record if an overhand knot is made, if 1 or 2 bows are used. Discontinue after 1½ minutes.

III Mobile Gross Motor Domain

<u>Domain description and limitations</u>. This domain includes the performance of gross motor behavior involving the movement of the entire body from one place to another distant place in space. The use of equipment in the performance of at least one skill should be included. At least two movements should not involve equipment. Examples of behaviors that fulfill these requirements and could be included in a representative sample are:

(1) jumping

(2) hopping

(3) crawling

(4) walking, e.g.,

(a) walking on a line,

(b) walking on a balance beam,

(5) climbing, e.g.,

(a) climbing stairs,

(b) climbing a ladder,

(6) "bike" riding, e.g.,

(a) tricycle riding,

(b) scooter riding,

A. Functional Area: Walking

1. Walking a balance beam.

a. Stationary stand.

Equipment: A balance beam 6 feet long, 4 inches wide, standing 5 inches off the floor.

Directions: Child is taken to the balance beam and examiner says: "In the next game we use this board. Let me see if you can stand on the board. Put one foot on the board. Good. Now put the other foot on it." If the child is able to place both feet on the board and maintains balance momentarily, proceed. If the child is unsuccessful, give another trial. If still unsuccessful, proceed to jumping.

b. Distance.

Examiner says, "Now let me see if you can walk all the way without falling off. Get back up on the board again and walk carefully on it." Record whether or not the child falls off. The examiner may at no time provide any physical assistance.

c. Speed.

Examiner says, "This time I want to see how fast you can walk the length of the board without falling off. Ready. Go." Begin timing as soon as one foot is lifted off the floor and stop timing as soon as one foot touches the floor. Once again the examiner should not provide any physical assistance. Record time only if child does not fall off.

B. Functional Area: Climbing

1. Stair climbing: Upstairs

<u>Equipment</u>: Standard stairs: 10 inches deep, 7 inches high.

<u>Directions</u>: The examiner should encourage the child to walk upstairs recording (1) whether or not railing or wall is used for assistance and (2) whether or not the child

placed both feet on each step (mark-time stepping), or alternates feet from one step to the next. At least 3 steps should be available. Examiner says, "Now I want you to walk up these stairs. Go ahead." While the examiner may walk beside the child to offer encouragement, s/he must <u>not</u> offer any physical assistance. If the child wants to take the examiner's hand, say "I want to see <u>you</u> do it.

2. Stair climbing: Downstairs.

<u>Equipment</u>: Standard stairs: 10 inches deep, 7 inches high.

<u>Directions</u>: Administer as in item 1 except ask the child to walk down the steps. The examiner should walk in front of the child to protect him/her from a fall. However, the examiner should not assist the child. Record: (1) if assistance (e.g., railing or wall) was used, and (2) if a mark-time stepping or an alternating stepping pattern was used.

C. Functional Area: Jumping

1. Both feet (broad jump).

<u>Equipment</u>: None

Directions: Place a 10 inch strip of masking tape on the floor. The examiner should first demonstrate the broad jump. That is, demonstrate jumping from a stationary position (both feet together) with toes touching the strip of tape. Ask the child to do the same. The examiner should measure the distance of the child's jump from the tape line to the child's heel position after the jump. Also record whether the child completed the jump with both feet together.

Examiner says: "I want to see how far you can jump with both feet together. Watch me do it first. I stand behind this line with my toes touching it and then I jump as far as I can with both feet together." Demonstrate jump. "Now I want you to jump as far as you can but keep both feet together. Stand here. When I say 'Go.' you jump as far as you can. Ready. Go."

2. Right foot (hop).

Equipment: None.

Directions: Using the tape line from item 1, the examiner should demonstrate jumping on the right foot. Measure and record the distance the child jumps before placing the other foot on the floor. Also record the number of completed jumps. The examiner should not physically assist the child in any way.

Examiner says: "Now I want to see how far you can jump on one foot. Watch me do it first." Examiner demonstrates. "Now you do it. When I say 'Go' lift this foot (touch left foot) and jump as far as you can go. Go as straight as you can and do not turn. Stand here. Ready. Go."

3. Left foot (hop).

Equipment: None.

Directions: Administer as item 2 except touch the child's right foot. It is not necessary for the examiner to demonstrate again.

Scoring Procedures

Rationale. Scoring of the Preschool Perceptual Motor Scale (PPMS) was designed for the experimental purposes of this study. When the final version of this test is developed, it should be administered to a large representative sample of three, four, and five year olds. The data gathered from this administration can then be used to determine allocation of points for quantitative and qualitative aspects of performance, and an individual's ranking on each motor skill. Since the overall size of the sample of children in this study was small, and the numbers of males and females within and across age groups were unequal, and geographic and socioeconomic limitations were present, temporary scoring procedures were established. Process and product scores were derived by comparing the normative data from the PPMS with similar data from the literature.

While normative data from the literature compensated for the limitations of the data from this study, it also presented restrictions. Specifically, at times, information was dated, incomplete or nonexistent. For example, since little is known about the various components of certain motor skills, some scores reflected only quantitative aspects of performance. Furthermore, when information about qualitative aspects of performance did exist, often information concerning the relative importance of each component was not specified. In addition, little information existed concerning the relative importance of any motor task to overall motor development. As a result, number of points within tasks, functional areas, and domains reflects more the availability of information than a deliberate attempt to use specific uniform guidelines. If both qualitative and quantitative information was available, product and process scores were used. However, if only one kind of information was available, only one kind of score was used. Therefore, task scores within functional areas could not be deliberately weighted to reflect relative importance. More information is needed before any specific guidelines are used for assigning points.

Working under these limitations, information from the literature and the normative data from the present study were compared to see if the developmental sequence for a particular motor skill, and the age levels at which the skill was accomplished were similar. When both sources of information agreed, the largest number of points was awarded to those aspects of performance which were shown to occur at later developmental levels or stages. However, assignment of points was more problematic when the data from this study did not agree with the normative information from the literature. In this case, judgements of the author were used. These judgements were based on the availability of the dates of the information, as well as the variety and completeness of the sources. While this information varied from one motor area to another, the same basic procedure was followed. To illustrate the process a detailed example will be provided. In addition, a general description will be given of the scoring processes for each of the other tasks in each domain. Finally, specific point allocations will be specified for each task.

Data Comparison Procedures: A Detailed Example

Under the stationary gross motor domain the scoring judgements for the two throwing tasks were made by the following process. Wild's (1938) description of the developmental stages in acquiring throwing skill was replicated. However, the children in the present study progressed through the developmental sequence approximately six months to a year earlier. Nevertheless, the younger the child the more likely s/he would stand facing the direction of the throw and toss the ball underhanded with little or no weight shift. The older the child, the more likely the rest of the body was incorporated so that eventually the child took a step and made a definite weight shift. Correspondingly, the children of the present study improved the distances of their throws as these refinements developed. Therefore, for the small ball throw, one point was given if the throw was overhand and no points if it was underhand. An additional point was awarded if the child stepped while throwing the ball. A third point was given if the child rotated the body by stepping with the foot opposite the side of the throw. These qualitative aspects of throwing had to be observed on two of three trials before the points were awarded. These points were summed to form a process score for each individual on the two throwing tasks.

Wellman (1937) reported the distances that 50 percent of children within various preschool age ranges were able to throw both large and small balls. The small ball she used was very comparable in size (9½ inches in circumference) to the one used in the present study (10 inches). However, the large ball (16¼ inches) used by Wellman was three and three-quarter inches smaller than the ball used in this study.

Wellman found children at two and one-half years of age throwing the small ball a distance of four to five feet. She reported that by three and one-half years throws of six to seven feet were common. The children who were four and one-half years of age in Wellman's study threw the small ball approximately 12 to 13 feet. The large ball proved more difficult to throw, and in general was thrown about two feet less at each age than was the smaller ball.

The average median distances across testing occasions for the throws of the three age groups of children in the present study were somewhat different. While 50 percent of the three year olds threw the small ball a very similar distance

(6½ feet) as did the three year olds in Wellman's study, the four year olds threw the ball approximately three feet less. In fact, the median distance of 11 feet accomplished by the five year olds was still a foot short of the four and one-half year olds in Wellman's study. While the distances for the large ball throw were comparably lower than Wellman's data, they did show a difference of approximately two feet less at each age than did the distances for the small ball throw.

While the applicability of Wellman's data was restricted by the ages of her subjects, the differently sized large ball, and the date when the data was collected, data from the present study also presented problems. The demographic limitations of the sample of children, as well as the wide variability in throwing performance restricted their direct application.

Considering the limitations of both sources of information, a compromised solution was considered necessary. The resulting scoring scheme reflected the fact that the same basic developmental sequence reported by Wellman was also present in the current study, but occurred at a later age. For the small ball throw, three points were given for any throw of 12 feet or more; two points were given for distances of six to 12 feet; and one point was given for throws below six feet. For the large ball throw, the same point distribution was repeated but for distances of two feet less for each age group.

Specifically, three points were given for distances of ten feet or less; two points for distances of four to ten feet; and one point was given for throws of less than four feet. These points comprised the product score for the throwing tasks. The process and product scores were then combined to provide a task score for each individual.

The same data comparison procedures were followed in allocating points for the functional areas of balancing, walking, catching and jumping. In all these cases, both product and process scores were awarded. The functional areas of writing and climbing had points awarded only for qualitative aspects of performance. However, the same procedures as were used for throwing tasks were used in these two domains.

The remaining three functional areas of kicking, copying, and dressing used only product scores. In the kicking and dressing functional areas the score directly reflected the number of accomplishments (e.g., number of kicks on three trials, number of buttons buttoned). Copying task scores reflected the number of details incorporated into the child's drawing. Task scores for each individual were summed to provide functional area, domain and total scores.

<u>Specific Allocation of Points</u>

I. Stationary Gross Motor Domain

A. Right Foot Balance (maximum Points: 4)

1. Process score
   - Use of assistance
     - 1 point if none used
     - 0 points if assistance used

2. Product score
   - 1 point if time less than 3 seconds
   - 2 points if time is 3 to 5 seconds
   - 3 points if time is greater than 5 seconds

B. Left Foot Balance
same as for right foot balance

C. Small Ball Throw (Maximum Points: 6)

1. Process score
   - 1 point for overhand throw
   - 1 point if steps while throwing
   - 1 point if steps with foot opposite side of throw

2. Product score
   - 3 points for distances of 12 feet or more
   - 2 points for distances of 6-12 feet
   - 1 point for distances of less than 6 feet

D. Large Ball Throw (Maximum Points: 5)

1. Process score
   - 1 point for overhand throw
   - 1 point for stepping with throw

2. Product score
   - 3 points for distances of 10 feet or more
   - 2 points for distances of 7-10 feet
   - 1 point for distances of less than 7 feet

E. Ground Level Kick (Maximum Points: 3)

1. Product score
   - number of kicks in a forward direction on three trials.

F. Drop Kick (Maximum Points: 4)

1. Process score
   - 1 point for coordinating release with kick of foot

2. Product score
   3 points for distances of 12 feet or more
   2 points for distances of 6-12 feet
   1 point for distances of less than 6 feet

D. Large Ball Throw (Maximum Points: 5)

1. Process score
   1 point for overhand throw
   1 point for stepping with throw

2. Product score
   3 points for distances of 10 feet or more
   2 points for distances of 7-10 feet
   1 point for distances of less than 7 feet

E. Ground Level Kick (Maximum Points: 3)

1. Product score
   number of kicks in a forward direction on three trials.

F. Drop Kick (Maximum Points: 4)

1. Process score
   1 point for coordinating release with kick of foot

2. Product score
   same as ground level kick

G. Throw-Catch (Maximum Points: 4)

1. Process score
   3 points for relaxed arm position
   2 points for stiff arm position
   1 point for cradle arm position

2. Product score
   1 point for 2 catches on 3 trials

H. Bounce Catch (Maximum Points: 3)

1. Product score
   number of catches on 3 trials.

2. Product score
same as ground level kick

G. Throw-Catch (Maximum Points: 4)

1. Process score
3 points for relaxed arm position
2 points for stiff arm position
1 point for cradle arm position

2. Product score
1 point for 2 catches on 3 trials

H. Bounce Catch (Maximum Points: 3)

1. Product score
number of catches on 3 trials.

II. Manipulative Small Motor Domain

A. Name Writing (Maximum Points: 7)

1. Process scores
Scribbles vs. letters
1 point for scribbles only
2 points for scribbles plus letters

Capital vs. lower case letters
1 point for all capital letters
2 points for mixture of capital and small letters

Reversals
1 point if no letter reversals
1 point if no word reversals

Spelling
1 point if correct spelling

B. Copying Cross (Maximum Points: 2)

1 point for intersecting lines (2 of 3 trials)
1 point for design orientation within 45 degrees of stimulus (2 of 3 trials)

C. Copying Circle (Maximum Points: 2)

1 point for circle figure vs. oval (2 of 3 trials)
1 point for closed figure (2 of 3 trials)

D. Copying Square (Maximum Points: 2)

1 point if 3 of 4 corners not rounded (2 of 3 trials)
1 point if 3 sides are equal (2 of 3 trials)

E. Copying Triangle (Maximum Points: 2)

1 point if 2 of 3 corners not rounded (2 of 3 trials)
1 point if design orientation within 45 degrees of stimulus (2 of 3 trials)

F. Copying Rectangle (Maximum Points: 3)

1 point if 3 of 4 corners not rounded (2 of 3 trials)
1 point if 2 sides greater than 2 other sides (2 of 3 trials)
1 point if drawing orientation within 45 degrees of stimulus (2 of 3 trials)

G. Copying Rectangular Design (Maximum Points: 5)

3 points same as copying rectangle
1 point if a line inside drawing is present (2 of 3 trials)
1 point if line drawn between lower left corner and upper right corner (2 of 3 trials)

H. Snapping (Maximum Points: 8)

1 point for each snap unsnapped
1 point for each snap snapped

I. Zipping (Maximum Points: 2)

1 point for zipping both up and down
1 point for latching zipper

J. Buttoning (Maximum Points: 8)

1 point for each button unbuttoned
1 point for each button buttoned

K. Lacing (Maximum Points: 10)

1 point for each eyelet laced

L. Tying (Maximum Points: 3)

1 point for tying overhand knot
1 point for tying 1 bow
1 point for tying 2 bows

III. Mobile Gross Motor Domain

A. Board Walking (Maximum Points: 4)

1 point if walked full length (1 of 2 trials)
3 points for walking time less than 5 seconds
2 points for walking time 5-7 seconds
1 point for walking time greater than 7 seconds

B. Climbing Upstairs (Maximum Points: 2)

1 point for using no assistance
1 point for alternating foot pattern.

C. Climbing Downstairs (Maximum Points: 2)

same as climbing upstairs

D. Broad Jumping (Maximum Points: 5)

1 point if both feet leave floor simultaneously
4 points if distance equal or greater than 40 inches
3 points if distance 20 to 29 inches
2 points if distance 13 to 19 inches
1 point if distance less than or equal to 12 inches

E. Left Foot Hop (Maximum Points: 4)

1 point if hops 2 or more hops
3 points if distance greater or equal to 120 inches
2 points if distance 40 to 119 inches
1 point if distance less than or equal to 39 inches

F. Right Foot Hop (Maximum Points: 4)

1 point if hops 2 or more hops
3 points if distance greater than or equal to 120 inches
2 points if distance 80 to 119 inches
1 point if distance less than 80 inches

APPENDIX C:

ANALYSES OF VARIANCE TABLES

Table 26. Analysis of variance table, regression coefficients for task score right foot lifted balance.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	114.34	22.87	9.94	0.0001	0.39
Error	77	177.17	2.30			
Corrected Total	82	291.52				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	64.03	13.91	0.0001
Sex	1	0.14	0.06	0.7985
Sex X Age	2	1.16	0.25	0.7816

Table 27. Analysis of variance table, regression coefficients for task score left foot lifted balance.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	30.64	6.13	7.44	0.0001	0.33
Error	77	63.39	0.82			
Corrected Total	82	94.02				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	14.18	8.61	0.0007
Sex	1	0.17	0.21	0.6509
Sex X Age	2	0.13	0.08	0.9212

Table 28. Analysis of variance table, regression coefficients for task score small ball throw.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	55.31	11.06	6.15	0.0002	0.29
Error	77	138.57	1.80			
Corrected Total	82	193.88				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	6.83	1.90	0.1547
Sex	1	13.77	7.65	0.0071
Sex X Age	2	2.25	0.62	0.5429

Table 29. Analysis of variance table, regression coefficients for task score large ball throw.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	25.52	5.10	4.90	0.0009	0.24
Error	77	80.17	1.04			
Corrected Total	82	105.69				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	0.10	0.05	0.9523
Sex	1	12.91	12.40	0.0011
Sex X Age	2	0.89	0.43	0.6605

Table 30. Analysis of variance table, regression coefficients for task score ground level kick.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	0.06	0.01	0.08	0.9927	0.00
Error	77	10.93	0.14			
Corrected Total	82	10.99				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	0.01	0.03	0.9732
Sex	1	0.01	0.04	0.8343
Sex X Age	2	0.01	0.03	0.9732

Table 31. Analysis of variance table, regression coefficients for task score drop kick.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	1.22	0.24	1.22	0.3067	0.07
Error	77	15.40	0.20			
Corrected Total	82	16.63				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	0.58	1.46	0.2382
Sex	1	0.01	0.06	0.8076
Sex X Age	2	0.11	0.28	0.7584

Table 32. Analysis of variance table, regression coefficients for task score throw catch.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	0.43	0.09	0.72	0.6112	0.04
Error	77	9.11	0.12			
Corrected Total	82	9.54				

* **

Source	DF	Partial SS	F Value	Prob F
Age	2	0.01	0.05	0.9550
Sex	1	0.11	0.95	0.6664
Sex X Age	2	0.12	0.49	0.6221

Table 33. Analysis of variance table, regression coefficients for task score bounce catch.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	3.82	0.76	4.76	0.0010	0.24
Error	77	12.35	0.16			
Corrected Total	82	16.17				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	3.09	9.65	0.0004
Sex	1	0.21	1.32	0.2534
Sex X Age	2	0.38	1.19	0.3091

Table 34. Analysis of variance table, regression coefficients for dependent variable writing.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	66.98	13.40	6.46	0.0001	0.29
Error	77	159.65	2.07			
Corrected Total	82	226.63				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	42.73	10.31	0.0003
Sex	1	0.89	0.43	0.5206
Sex X Age	2	0.08	0.02	0.9819

Table 35. Analysis of variance table, regression coefficients for task score copying a cross.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	2.27	0.45	1.69	0.1458	0.10
Error	77	20.65	0.27			
Corrected Total	82	22.93				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	1.10	2.05	0.1341
Sex	1	0.00	0.02	0.8942
Sex X Age	2	0.34	0.64	0.5339

Table 36. Analysis of variance table, regression coefficients for task score copying a circle.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	0.51	0.10	0.81	0.5467	0.05
Error	77	9.75	0.13			
Corrected Total	82	10.27				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	0.29	1.16	0.3179
Sex	1	0.28	2.23	0.1353
Sex X Age	2	0.14	0.57	0.5723

Table 37. Analysis of variance table, regression coefficients for task score copying a square.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	20.51	4.10	10.13	0.0001	0.40
Error	77	31.18	0.40			
Corrected Total	82	51.69				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	12.58	15.53	0.0001
Sex	1	0.20	0.48	0.5036
Sex X Age	2	0.21	0.26	0.7764

Table 38. Analysis of variance table, regression coefficients for task score copying a triangle.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	22.68	4.54	7.08	0.0001	0.31
Error	77	49.35	0.64			
Corrected Total	82	72.02				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	14.20	11.07	0.0002
Sex	1	0.00	0.00	0.9633
Sex X Age	2	1.50	1.17	0.3169

Table 39. Analysis of variance table, regression coefficients for task score copying a rectangle.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	53.19	10.64	10.91	0.0001	0.41
Error	77	75.08	0.98			
Corrected Total	82	128.27				

* **

Source	DF	Partial SS	F Value	Prob F
Age	2	22.60	11.59	0.0001
Sex	1	0.59	0.60	0.5545
Sex X Age	2	1.97	1.01	0.3708

Table 40. Analysis of variance table, regression coefficients for task score copying a rectangular design.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	131.42	26.28	13.04	0.46	0.46
Error	77	155.18	2.02			
Corrected Total	82	286.60				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	78.32	19.43	0.0001
Sex	1	0.02	0.01	0.9158
Sex X Age	2	4.43	1.10	0.3388

Table 41. Analysis of variance table, regression coefficients for task score snapping.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	1.73	0.35	5.71	0.0003	0.27
Error	77	4.68	0.06			
Corrected Total	82	6.41				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	1.09	8.96	0.0006
Sex	1	0.01	0.20	0.6566
Sex X Age	2	0.05	0.38	0.6936

Table 42. Analysis of variance table, regression coefficients for task score zipping.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	3.78	0.76	3.61	0.0057	0.19
Error	77	16.10	0.21			
Corrected Total	82	19.88				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	2.92	6.99	0.0020
Sex	1	0.48	2.31	0.1290
Sex X Age	2	0.21	0.51	0.6074

Table 43. Analysis of variance table, regression coefficients for task score buttoning.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	11.15	2.23	8.00	0.0001	0.34
Error	77	21.47	0.28			
Corrected Total	82	32.63				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	10.15	18.20	0.0001
Sex	1	2.20	7.89	0.0064
Sex X Age	2	2.38	4.28	0.0170

Table 44. Analysis of variance table, regression coefficients for task score lacing.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	4.86	0.97	4.75	0.0010	0.24
Error	77	15.74	0.20			
Corrected Total	82	20.60				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	3.48	8.51	0.0007
Sex	1	0.31	1.50	0.2225
Sex X Age	2	0.81	1.97	0.1439

Table 45. Analysis of variance table, regression coefficients for task score tying.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	43.76	8.75	8.77	0.0001	0.36
Error	77	76.87	1.00			
Corrected Total	82	120.63				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	30.96	15.51	0.0001
Sex	1	3.77	3.77	0.0527
Sex X Age	2	0.79	0.40	0.6800

Table 46. Analysis of variance table, regression coefficients for dependent variable board walk.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	8.92	1.78	5.67	0.0003	0.27
Error	77	24.23	0.31			
Corrected Total	82	33.16				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	2.29	3.65	0.0298
Sex	1	0.44	1.41	0.2370
Sex X Age	2	0.69	1.09	0.3422

Table 47. Analysis of variance table, regression coefficients for task score climbing upstairs.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	5.20	1.04	2.79	0.0224	0.15
Error	77	28.68	0.37			
Corrected Total	82	33.88				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	0.99	1.33	0.2691
Sex	1	0.40	1.07	0.3041
Sex X Age	2	0.47	0.63	0.5415

Table 48. Analysis of variance table, regression coefficients for task score climbing downstairs.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	6.96	1.39	2.37	0.0464	0.13
Error	77	45.21	0.59			
Corrected Total	82	52.17				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	0.81	0.69	0.5092
Sex	1	1.99	3.38	0.0663
Sex X Age	2	0.27	0.23	0.7970

Table 49. Analysis of variance table, regression coefficients for task score broadjump.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	33.57	6.71	6.80	0.0001	0.31
Error	77	75.99	0.99			
Corrected Total	82	109.57				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	21.39	10.83	0.0002
Sex	1	0.11	0.11	0.7392
Sex X Age	2	0.60	0.30	0.7427

Table 50. Analysis of variance table, regression coefficients for task score left foot jump.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	46.89	9.37	7.71	0.0001	0.33
Error	77	93.71	1.22			
Corrected Total	82	140.60				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	23.20	9.53	0.0004
Sex	1	0.06	0.05	0.8164
Sex X Age	2	0.57	0.23	0.7951

Table 51. Analysis of variance table, regression coefficients for task score right foot jump.

Source	DF	SS	MS	F	Prob F	R^2
Regression	5	47.42	9.48	7.07	0.0001	0.31
Error	77	103.32	1.34			
Corrected Total	82	150.75				

* *

Source	DF	Partial SS	F Value	Prob F
Age	2	26.53	9.89	0.0003
Sex	1	0.13	0.10	0.7536
Sex X Age	2	0.17	0.06	0.9376

ACKNOWLEDGMENTS

I am pleased to acknowledge the dedication, and assistance of my wife, Shirley, without whose love and devotion this dissertation would not be possible. While I appreciate the help of my major professor, Gary Phye, and the members of my committee, their contributions are shadowed by the magnanimous contributions of my wife.